12-2018

Digging deeper: Development and application of an untargeted exometabolomics approach to identify biogeochemical hotspots of dissolved organic matter vulnerability in Arctic soils

Mallory Paige Ladd
University of Tennessee, mladd5@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Mallory Paige Ladd entitled "Digging deeper: Development and application of an untargeted exometabolomics approach to identify biogeochemical hotspots of dissolved organic matter vulnerability in Arctic soils." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Robert L. Hettich, Major Professor

We have read this dissertation and recommend its acceptance:

Shawn R. Campagna, Colleen M. Iversen, Sean M. Schaeffer, Stan D. Wullschleger

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Digging deeper: Development and application of an untargeted exometabolomics approach to identify biogeochemical hotspots of dissolved organic matter vulnerability in Arctic soils

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Mallory Paige Ladd
December 2018

DEDICATION

I dedicate this dissertation to every little girl who was told science and math are for the boys. And to every grown woman who has had to work twice as hard to prove that we belong here too.

ACKNOWLEDGMENTS

Completing this dissertation has been one of my most challenging academic pursuits. I would like to express my deepest gratitude to the following people who have supported me on this road to the PhD.

To my adviser Dr. Robert Hettich, thank you for opening your lab to a diverse group of graduate students and for maintaining a collaborative and supportive atmosphere where we could develop not only a sound analytical training but a firm foundation in experimental design, project management, and science communication, ensuring we become successful independent researchers.

I would like to thank my committee, Drs. Stan Wullschleger, Colleen Iversen, Shawn Campagna, and Sean Schaeffer, for the time and resources they have invested in me, for providing me the opportunity to get out of the lab and into the field to experience the complexity of the natural environment and collect my own samples, and for invaluable feedback on experimental design and guidance in manuscript preparation.

Also, I want to express my gratitude to the North Slope Borough and the Umiaq Science Support team in Utqiaġvik, and the Council Native Corporation, for providing logistical assistance, safety training, and for allowing us to visit, study, and work on Iñupiat lands. Thank you to my collaborators Neslihan Tas and Jana Phillips for assistance with experimental design and technical support. Huge thanks go out to my lab mates over the years; Paul Abraham, Rich Giannone, Weili Xiong, Sarvesh Iyer, Chen Qian, Lauren Swientoniewski, Suresh Poudel, Hannah Simpson, Manuel Ivan Villalobos Solis, J. Alfredo Blakeley-Ruiz, Alex Cope, David Reeves, Samantha Peters, Payal Chirania, and Cheng Chen for helping me dig in the dirt, run samples or ANOVAs, go on nature walks, or just being there in the hallway or at group meeting to provide valuable feedback on an upcoming presentation.

I am grateful to Dr. Lee Riedinger, Dr. Mike Simpson, Wanda Davis, and everyone in the Bredesen Center for not only creating the pioneering Energy Science and Engineering program but an academic atmosphere that truly became like a second family to me, something I believe is critical to our mental health on this PhD journey. I feel very

lucky to have made several lasting friendships from this group of amazing people. I cherish my female friendships most and there are far too many to thank them all individually. But to Christine Ajinjeru, Jayde Aufrecht, and Amber Hall Nelson I am eternally grateful. Whether it was to celebrate our successes with a night out or a flight overseas, or to wash away our frustrations and failures with a glass of wine and an evening of laughter, I knew I could always count on you two for sound advice and a dose of inspiration and perspective.

This work would not have been possible without the unwavering love, patience, guidance, and encouragement from my family. Thank you to my mom and dad, Kelly and Michael Ladd, and my Grandma Joanne and Grandma Mimi, for believing in me from the very beginning, and every day since. To my sister Kenz and my brother Zach for being such amazing role models and for helping me feel okay about the fact I was still in school at age 30. And to Tony Bova, my closest friend and partner in life. I will forever be grateful that I had you by my side through this all. Thank you for being a thoughtful sounding board for all my half-baked ideas, a grounding but inspiring force on the toughest of days, and for always knowing how to put a smile on my face. I love you with all my being.

Also, I would be remiss to leave out Folgers and Tazo, red wine and birth control. Y'all have done me a solid throughout this #PhDlife and I am so very, very, very thankful.

Last, but certainly not least, to Ellen DeGeneres. You didn't know this, but over these past five years, no matter how hard a day it was in lab, no matter who got elected, appointed, or confirmed, and no matter how many dollars were (or were not) in my bank account, it was your show that I would turn to for a hearty laugh and a daily reminder that there are more important things in life, that there is good in the world, and that all we need to do to add to that good is *be kind to one another*. Thank you for all that you do.

ABSTRACT

Arctic soils contain vast reserves of carbon (C) that, with rising temperatures, may become a significant source of greenhouse gases (GHGs) (*i.e.* CO_2, CH_4, N_2O) due to increased microbial decomposition of soil organic matter (SOM). However, there are significant spatial variations in GHG production that lead to hotspots of C release across the landscape, creating significant uncertainty in climate models. Reliably predicting the magnitude of C loss via microbial production of GHGs, and the proportion lost as either CO_2 or CH_4, depends on many factors, including soil temperature and moisture, microbial community structure and function, as well as the composition and availability of the most labile SOM pool—low molecular weight dissolved organic matter (LMW DOM). While the effects of temperature and moisture on GHG production in Arctic soils have been studied extensively, there is a dearth of information on the effects of LMW DOM chemistry and its potential to be a predictive chemical signal of biological hotspots of C release, in large part due to unique analytical challenges. LMW DOM is an incredibly complex and dynamic mixture of small molecules from both biotic and abiotic origin that turnover on the order of days or even hours and are obscured by countless other interfering signals in the soil, each a complicating factor in isolation, detection, and quantitation. Recent advancements in liquid chromatography mass spectrometry (LC/MS) have provided a means for sensitive, robust, and high-throughput measurements of LMW DOM composition and availability but have not yet been applied in Arctic soils. In this dissertation, an untargeted LC/MS approach for characterizing LMW DOM availability was developed and evaluated, benchmarking its analytical performance in Arctic soils for the first time. The optimized approach was then applied to soils from two Arctic ecosystems to measure variations in LMW DOM across the landscape, due to soil depth, aboveground vegetation, topography, or level of degradation due to thaw. In addition to establishing the LC/MS measurements and data interpretation, this dissertation also had several key interdisciplinary components including remote-location field sample collection, establishing an accessible data analysis pipeline, and examining this work from a public policy perspective.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

6-MAP	6-methylaminopurine riboside
ACN	acetonitrile
AI	aromaticity index
ANILCA	Alaska National Interest Lands Conservation Act
ANOVA	analysis of variance
ARPA	Arctic Research and Policy Act
BEO	Barrow Environmental Observatory
CH_4	methane gas
CIA	compound identification algorithm
CID	collision-induced dissociation
CO_2	carbon dioxide gas
CUE	carbon use efficiency
CV	coefficient of variance
DBE	double bond equivalents
DDA	data-dependent acquisition
DNA	deoxyribonucleic acid
DOE	U.S. Department of Energy
DOM	dissolved organic matter
DTLB	drained thaw lake basins
EEZ	exclusive economic zone
ESI	electrospray ionization
FA	formic acid
FC	fold change
FDR	false discovery rate
FT-ICR	Fourier-transform ion cyclotron resonance
FWHM	full-width half-maximum
GC	gas chromatography
GHG	greenhouse gas
GUI	graphical user interface
HCD	high-energy collisional dissociation
HCP	high-centered polygon
HILIC	hydrophilic interaction chromatography
HQFs	high-quality features
HRMS	high-resolution mass spectrometry
IC	inorganic carbon
IPA	isopropyl alcohol
LC	liquid chromatography
LCP	low-centered polygon
LMW	low molecular weight
LOESS	locally estimated scatterplot smoothing
LTQ	linear trap quadrupole
m/z	mass-to-charge ratio

MeOH	methanol
MRT	mean residence time
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS^1	molecular ion
MS^2	fragment ion
N_2O	nitrous oxide
NASA	National Aeronautics and Space Administration
NDIR	nondispersive infrared
NGEE	Next-Generation Ecosystems Experiments
NH_4	ammonium
NH_4Ac	ammonium acetate
NMR	nuclear magnetic resonance
NO	nitrogen monoxide
NO_3	nitrate
NSPD	National Security Presidential Directive
NUE	nitrogen use efficiency
O_3	ozone
ORNL	Oak Ridge National Laboratory
PCA	principal component analysis
PLS-DA	partial-least squares discriminant analysis
ppm	parts per million
QC	quality control
QIT	quadrupole ion trap
QQQ	triple quadrupole
RANSAC	random sample consensus algorithm
RNA	ribonucleic acid
RP	reversed-phase
RT	retention time
SOA	secondary organic aerosols
SOC	soil organic carbon
SOM	soil organic matter
S/N	signal-to-noise ratio
TC	total carbon
TOC	total organic carbon
TOF	time-of-flight
UNCLOS	United Nations Law of the Sea Convention
UV-Vis	ultra violet-visible
XIC	extracted ion chromatogram

CHAPTER 1: PRINCIPLES OF MASS SPECTROMETRY-BASED EXOMETABOLOMICS AND APPLICATIONS TO CLIMATE SCIENCE

Parts of this chapter have been adapted from the following manuscripts:

Ladd, M.P., Abraham, P., Giannone, R., Hettich R. Evaluation of an untargeted nano-liquid chromatography, dual-polarity, tandem mass spectrometry approach to expand coverage of low molecular weight dissolved organic matter in Arctic soil. *Scientific Reports (in review).*

Ladd, M.P., Reeves, D., Poudel, S., Iversen, C.M., Wullschleger, S.D. Hettich, R.L. Untargeted exometabolomics reveals biogeochemical hotspots with vegetation and polygon type in arctic tundra soils. *Environmental Science & Technology* (*in prep*).

ML's contributions included: literature review, experimental design, sample collection, sample preparation, data collection, data analysis, manuscript writing and editing.

1.1   Climate change: a defining *energy* challenge of this generation

Throughout history, humans have consistently found new and better sources of energy to enhance our abilities; whether that be our ability to communicate with one another, to move from place to place, to make our homes brighter or warmer, or to lift, lower, push, pull, or turn something faster or further. Early on, we burned wood to keep warm, provide light, and prepare food. During early industrial development, we added wind and hydropower, and around the mid- to late-1800s, with an ever-growing need for better tools, transportation, and electricity, coal became our primary source of energy. Two other fossil fuels, oil and natural gas, were quick to follow, rounding out the three major sources that now supply nearly 80 % of the world's energy (Figure 1).[1] It was the burning of fossil fuels that enabled human civilization to grow to unimaginable heights. However, it was also these combustion processes that released more and more carbon dioxide (CO_2) into the atmosphere (Figure 2), where it acted as a greenhouse gas (GHG), absorbing and trapping heat in the Earth's atmosphere. Consequently, alongside rising CO_2 levels, the planet also started to experience warmer temperatures with each passing year (Figure 3).
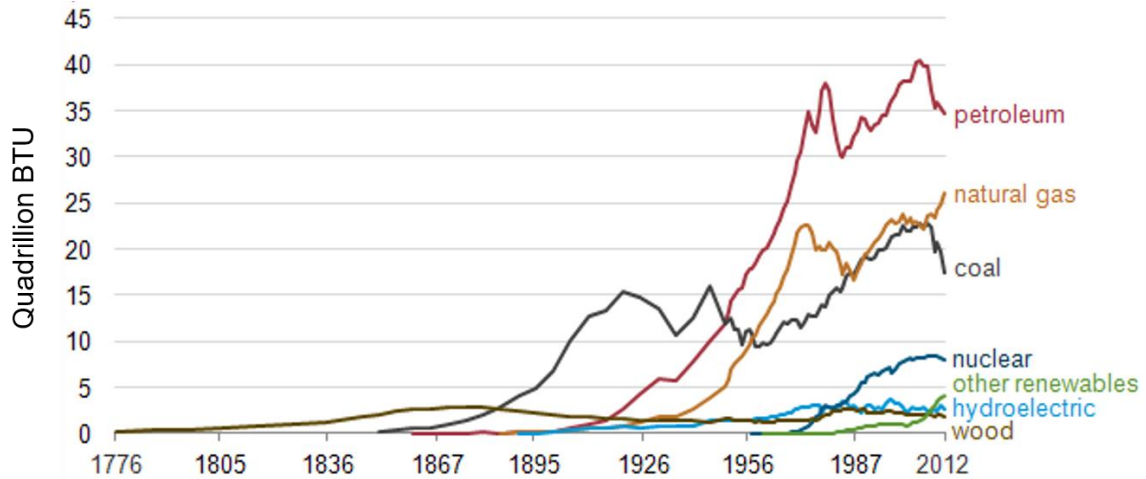
Figure 1: History of energy consumption in the United States (1776-2012)

Source: Public Domain, U.S. Energy Information Administration[1]
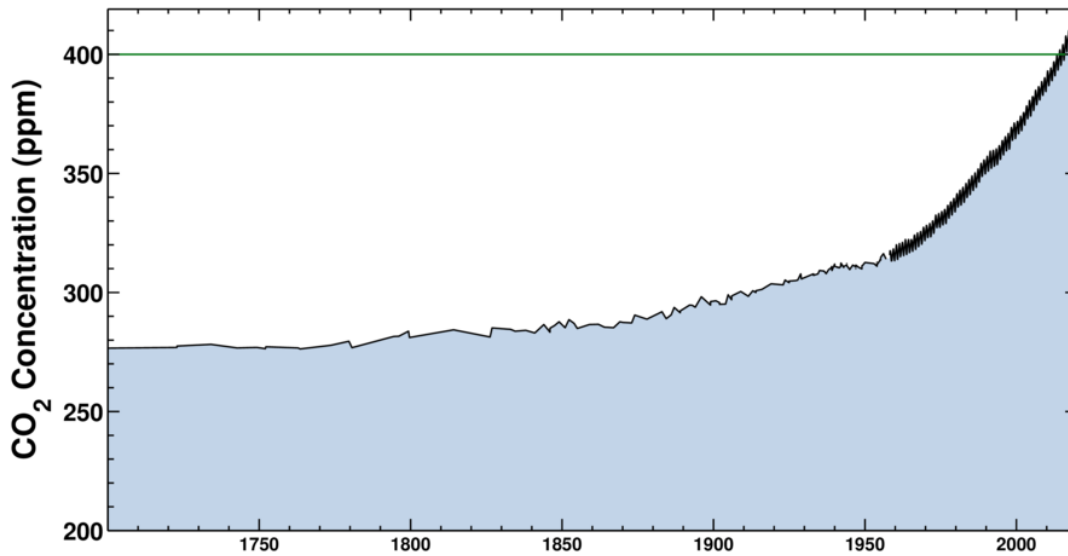
Figure 2: Global atmospheric CO_2 concentrations (1700-present)

"The Keeling Curve." Data obtained from ice cores prior to 1958 and from the Mauna Loa Observatory after 1958. Green line indicates most recent reading from September 18, 2018 at 405.83 ppm. Source: Public Domain, Scripps Institution of Oceanography[2]
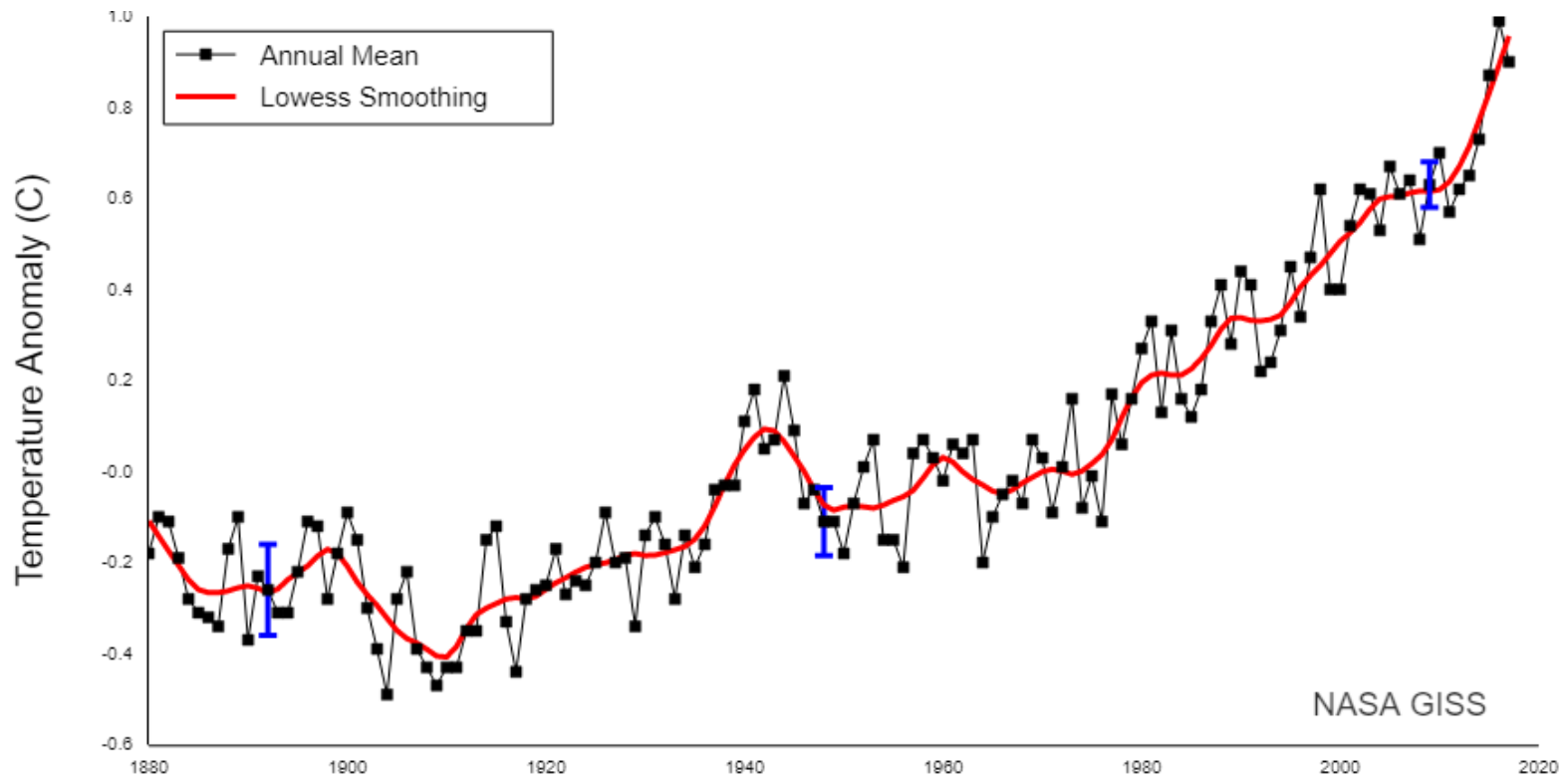
Figure 3: Global mean temperature anomaly estimates based on land and ocean data (1880-present)

Source: Public Domain, National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies[3]

While climate change has sometimes been characterized as solely an environmental issue, like air or water pollution for example—erroneously suggesting that it may be addressed by simply cleaning up a few bad habits—it has become one of the most pressing and complex *energy* challenges of this generation. Warmer temperatures, rising sea levels, and increased instances of severe weather will have far-reaching effects not only on the environment, human health, and national security, but also on how we grow our food, how we move goods and do business, and how we extract, generate, transport, and use energy resources (Figure 4).[4] For example, changes in water availability due to drought will impact our ability to cool power plants, generate hydroelectric power, or grow biofuel feedstocks. Higher temperatures in the summer or lower temperatures in the winter will impact how we heat and cool our homes and businesses, subsequently altering electricity demands, requiring new infrastructure and technologies for distribution and storage.[5] Thus, our ability to make informed decisions about how to manage our energy generation and use in the future depends heavily on our understanding of, and ability to predict climate change.

Global predictions of climate rely on computational models and data collected at finer scales, at the regional or landscape level for example, all the way down to biogeochemical processes occurring at the molecular scale.[6] Each of these models has multiple variables, feedbacks between processes or scales, and varying levels of detail—spatial/temporal resolution—and uncertainty.[7] Reducing this uncertainty enables scientists and policymakers alike to make more informed decisions about future research directions and climate or energy policy agendas. One geographical area of considerable uncertainty, why it is often referred to as an "adaptation tipping point" for climate change, is the Arctic.

5
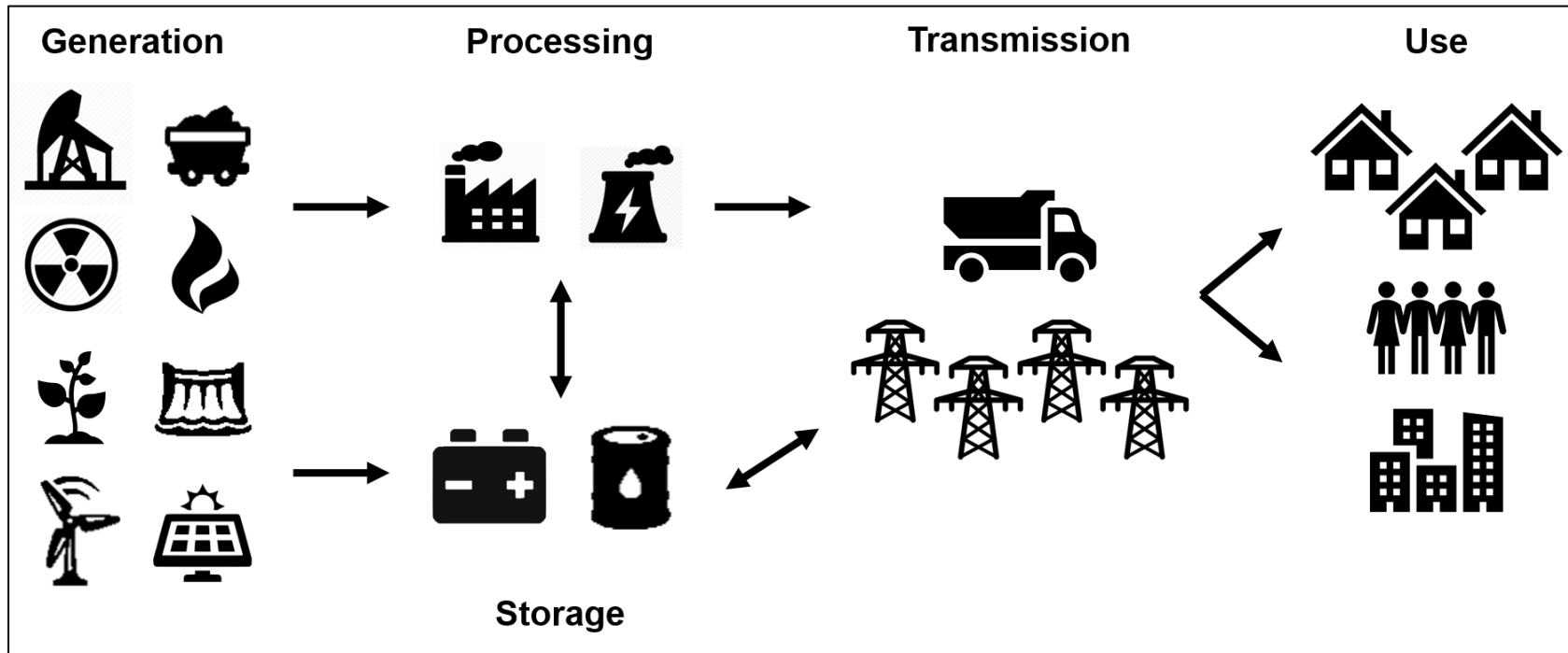
Figure 4: Illustration showing the various energy sources and steps—from generation to use by a household or business—that will be impacted by climate change

1.2 Why the Arctic: unique and sensitive to change

Historically characterized for its remote, boundless, snow-white landscapes and pristine beauty, more recently the Arctic has become a synonymous symbol for climate change. Defined as the area north of 66 °N latitude or the area north of the tree line where permanently frozen ground (permafrost) becomes continuous across the landscape, the Arctic is also known for its unique radiative cycles that, after spring snowmelt during the summer months, have earned it the nickname "land of the midnight sun." Following the short and cool growing seasons however, the landscape is promptly covered in snow again, and the frigid temperatures and long dark winter months return (Figure 5). Accordingly, any plants or animals that cannot survive the winter become a part of the frozen landscape, slowly decomposing into soil organic matter (SOM) rich in carbon (C) and other nutrients. With this cycle repeating each year for millennia, the Arctic has traditionally acted as a carbon "sink," now storing nearly half the Earth's terrestrial C stocks in SOM associated with permafrost soils.[8, 9]

However, the Arctic is also warming twice as fast as any other landscape on the planet.[10] Rising temperatures have accelerated permafrost thaw, both in depth and duration, resulting in physical, hydrological, and chemical shifts across the landscape, leading to previously-frozen SOM suddenly becoming available for microbial decomposition.[11-13] Mobilizing even a fraction of this C-rich SOM via these geomorphological and biochemical processes is projected to increase the release of GHGs like CO_2, methane (CH_4), and nitrous oxide (N_2O) from the landscape, creating a significant positive feedback to climate change.[14-17]

7
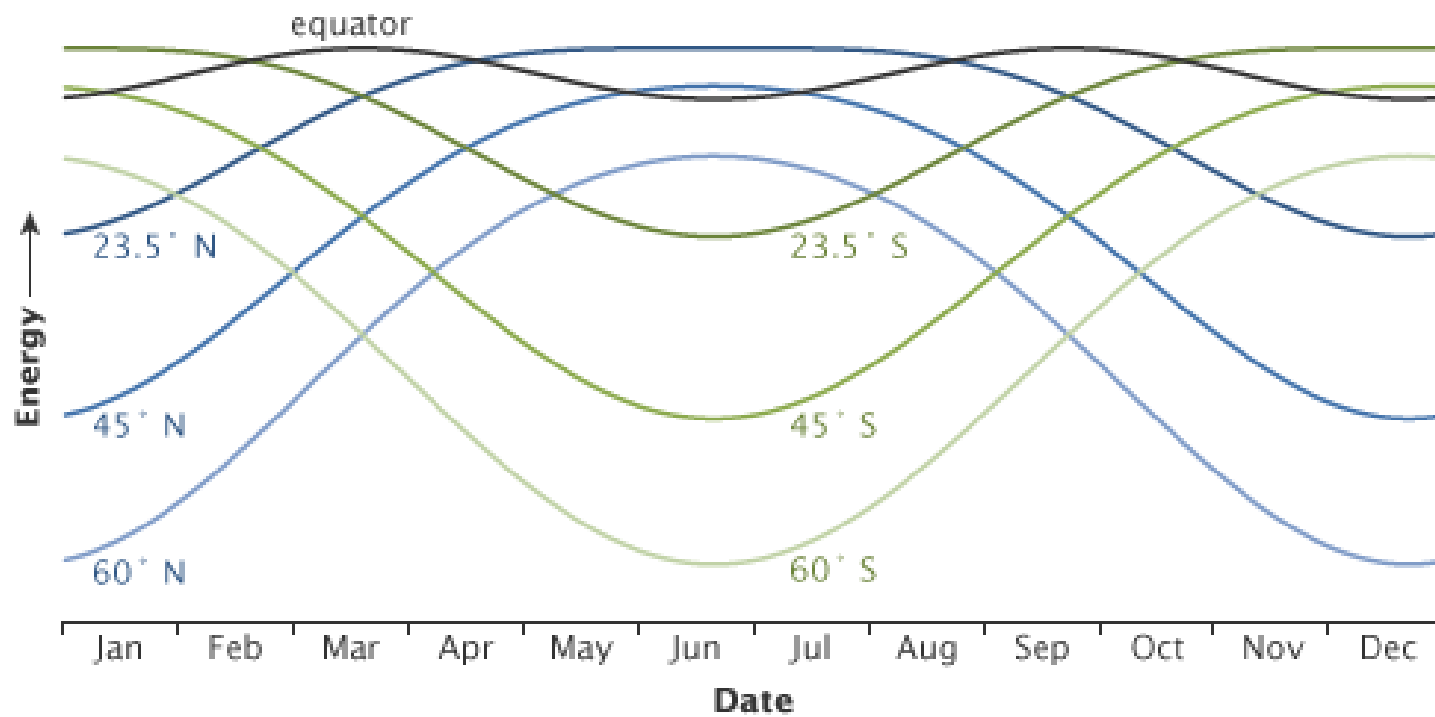
Figure 5: Solar radiation across Earth's latitudes over the course of one calendar year

Higher radiative energy is observed during the summer months and lower energy during the winter months at higher latitudes like the Arctic. Source: Public Domain, NASA Earth Observatory.[18]

Despite this, the underlying mechanisms regulating GHG release are not well defined, as they can vary both temporally and spatially and are impacted by both biotic and abiotic variation. For example, in addition to warming temperatures directly increasing microbial metabolism, it may also increase plant productivity, particularly in shrubs,[19] which could act as a mitigating *negative* feedback due to higher photosynthesis rates and root exudation leading to C sequestration in the plant biomass and belowground.[20] In contrast, increased root exudation may also stimulate SOM turnover, a process called the *priming effect*.[21] Another complicating factor is that along with enhanced C mobilization, organic nitrogen (and phosphorus) may also be released from thawing permafrost soils impacting both plant and microbial community activity.

Nitrogen (N) is essential to all organisms but is generally limiting in terrestrial ecosystems,[22] which results in competition between the plant and microbial communities.[23, 24] Most of the N input into soils is from plant and microbial residues in the form of polymers (*i.e.* proteins, chitin, peptidoglycan). Microbial extracellular enzymes— which require C, N, and energy for their synthesis and expression—break those polymers into smaller, monomeric units where they can then be taken up directly by a plant[25-28] or microbe, the "direct route", or further degraded into mineral forms such as ammonium (NH_4) and nitrate (NO_3), the "mineralization-immobilization-turnover" (MIT) route (Figure 6).[29] Because N is limiting, microbes tightly regulate the synthesis and activity of extracellular enzymes according to the availability of substrates and their resource requirements (*i.e.* C:N ratio, carbon or nitrogen use efficiency, CUE, NUE), and generally prefer inorganic N sources.[30] However, under C-limiting conditions, in aerated systems
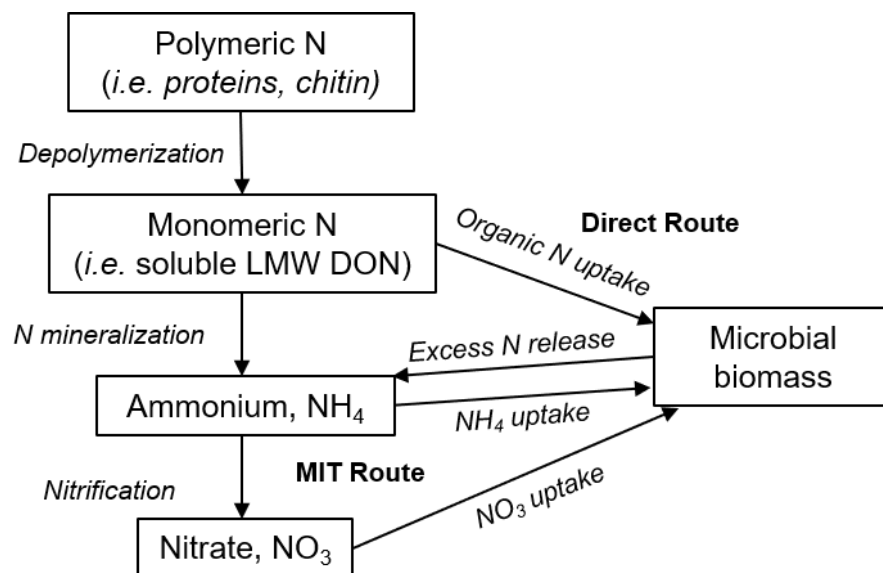
Figure 6: Typical microbial N utilization pathways in soils, from organic N input to incorporation into the microbial biomass via two competing routes

where NH_4 is low due to plant uptake, or when C:N ratios are high resulting in net N immobilization, the direct route becomes favored due to microbial communities using organic molecules as C sources.[29] Phosphorus (P) is also an essential element for life. Although it is involved in the synthesis of many key biomolecules including DNA, RNA, and ATP, it is primarily derived from the weathering of the parent rock material and thus, is also limiting across most terrestrial systems. Because of this, plants and microbes have evolved to have several acquisition strategies and can assimilate P in multiple forms (*i.e.* oxidation states). In the Arctic, due to the low temperatures and high moisture, organic P is the main source of plant and microbial P, some proportion of which is presumably found in the LMW DOM pool; however, only a handful of studies have been done to evaluate the molecular composition of organic P in Arctic soils.[31, 32]

In addition to the above- and belowground variability in plant and microbial resource requirements, the stoichiometry of SOM compounds, and the availability of those substrates each impacting the C balance in terrestrial ecosystems,[33, 34] each of these is also impacted by hydrology and changing seasonality in Arctic systems—longer growing seasons due to earlier spring snowmelt.[35, 36] For example, early spring is an important time for biogeochemical cycling due to snowmelt, which releases a flush of nutrients into the soil when the microbes are just starting to "wake up," but before the plants start to grow and compete for nutrients. The absence of snow also leads to more dynamic freeze-thaw cycles, which have been shown to impact microbial community structure and function, as well as C, N, and P availability.[37-41] In addition the Arctic becoming warmer, the melting of ice (which is prevalent in Arctic soils) and altered precipitation regimes are expected to cause the Arctic to become wetter as well. Changes in water availability, both spatially and temporally, will impact microbial community composition and activity (*i.e.* anaerobic conditions favoring methanogenesis) and the availability of SOM and nutrients, ultimately impacting the composition and magnitude of GHG release in Arctic soils.[42]

Thus, being able to reliably predict where hotspots (*i.e.* increased C-loss) are most likely to occur requires a detailed understanding of the relationship between landscape heterogeneity[43-45] and the associated shifts not only in hydrology (topography),[46, 47] vegetation,[48, 49] and microbial community composition,[50] but in the chemical composition of SOM, and its inherent *availability* to soil microbial communities[51-53]—molecular-scale information that is currently poorly understood and/or poorly characterized in process-based models.[54-56]

1.3   Analytical challenge: characterizing LMW DOM

At the molecular level, soil organic matter (SOM) is described as a continuum of progressively-decomposing organic material, whose composition is impacted by the turnover time of each pool and its availability (*i.e.* adsorption to mineral surfaces).[57] Historically, SOM turnover has generally been described at the bulk level, by the mean residence time (MRT), or half-life ($T_{1/2}$) using first-order modeling (Equation 1), or by measuring isotopic abundances (*i.e.* ^{13}C natural abundances, ^{14}C dating).[58]

$$\text{Equation 1: } \frac{\partial S}{\partial t} = I - kS,$$

where, S is the SOM stock, t is the time, k is the decomposition rate, and kS is equivalent to input, I. The MRT can then be calculated using Equation 2, and the $T_{1/2}$ by Equation 3.

$$\text{Equation 2: } MRT = \frac{1}{k}$$

$$\text{Equation 3: } MRT = T_{\frac{1}{2}}/ln2$$

Not all SOM degrades at the same rate however. The fraction of SOM most available to microbial decomposers, and thus most susceptible to mineralization and release, is the water-soluble fraction dominated by small organic molecules (< 1000 Da) found suspended in soil pore and surface waters—low molecular weight (LMW) dissolved organic matter (DOM). An incredibly complex and dynamic mixture, LMW DOM originates from and feeds back to both biotic and abiotic processes (*i.e.* plant root exudates, plant nutrients, products/substrates of microbial metabolism or turnover, photodegradation products), acting as both a reflection of and a control on biogeochemical cycling.[57, 59, 60] To

give some context on the complexity of this analyte pool and the processes acting upon it, it has been estimated that there can be upwards of 10^{10} microbial cells found in a single gram of soil.[61] Each of those cells is simultaneously taking up substrates and releasing byproducts of metabolism into the soil environment, where those compounds may then, for example, be taken up by a plant or another microbe, bind to a mineral surface, or degrade abiotically, depending on the conditions (see discussion above). These turnover processes involving LMW DOM have been shown to occur on the order of days, hours, or even minutes (20-40 minutes in a grassland soil[62]) under different conditions.[63] In laboratory incubations of Arctic soils, LMW DOM composition and turnover has been shown to be sensitive to variations in both temperature[63-65] and moisture.[66] Analogously, the structure and function of soil microbial communities are both strongly influenced by the molecular composition of this highly-labile substrate pool.[67, 68] This relationship between environmental conditions, plant and microbial communities, and LMW DOM composition ultimately determines how an ecosystem will respond under a changing climate. Despite it representing an information-rich chemical fingerprint of biological function in soil, and thus a potential indicator of SOM vulnerability that could help reduce uncertainty in process-based predictive models of C cycling,[55, 56] the molecular variability of LMW DOM across Arctic landscapes is largely unknown.

This is due in large part to unique analytical challenges that exist with soil matrices, including the wide-ranging physicochemical properties of LMW DOM, high rates of uptake and release of those analytes leading to consistently low concentrations, and the abundance of potentially interfering inorganic (*i.e.* salt) species, all of which pose

13

significant obstacles in isolation, detection, and quantitation.[69, 70] As such, most analyses of LMW DOM in Arctic soil have been at the bulk level (*i.e.* total organic carbon or nitrogen, separation by physical fractionation or solubility, colorimetric/fluorometric assays) or have targeted a specific subset of compounds—mainly, amino acids.[71-73] These bulk analyses often require the soil to be removed from its natural state and involve pretreatment steps that physically or chemically alter the composition before detection and quantitation, introducing bias or failing to elucidate complex interactions occurring at the microsite scale.[74, 75] In addition, although a valuable technique to quantify pools and fluxes, isotopic labeling studies often don't use ecologically-relevant concentrations to track the movement of organic monomers through the soil due to the insufficient analytical detection limits of established techniques, and generally targeted only a single compound or a small class of compounds.[27, 76]

Beginning instead with an untargeted approach however allows for the identification of biogeochemical hotspots and the generation of unbiased hypotheses about the biological functioning of these compounds under contrasting environmental conditions. For example, characterizing LMW DOM compounds and variation in their relative abundances over space, time, or under a perturbed environmental condition could help identify diurnal cycles of biological activity, distinguish rate-limiting steps in decomposition, monitor plant-microbial competition for organic nutrients, or, as in the case of this work, elucidate the controls on LMW DOM degradability and susceptibility to release as a GHG. After identifying ecologically-relevant metabolites or other small molecules that undergo a significant fold change (FC) between conditions, one could then

transition to a targeted analysis, improving specificity and allowing for additional experiments to be carried out (*i.e.* absolute quantitation, flux analyses). Ultimately, this kind of comprehensive molecular knowledge has significant potential to provide novel insights into microbially-mediated processes in soil and offer an improved fundamental understanding of C and/or N cycling in the Arctic.

Because of this, the field has increasingly been turning to untargeted approaches to characterize organic matter in Arctic soils using a variety of techniques including nuclear magnetic resonance (NMR) spectroscopy,[77] ultraviolet-visible (UV-Vis) or excitation-emission matrix fluorescence spectroscopy,[78] or gas chromatography/mass spectrometry (GC/MS).[79, 80] However, due to inherent limitations associated with these techniques, including inadequate detection sensitivity, limited dynamic range, or a need for chemical derivatization prior to analysis, there has been increased interest in evaluating mass spectrometry-based approaches that offer higher sensitivity and both qualitative and quantitative information within one analysis.[81-83] In particular, because of recent advances in instrumentation and informatics tools, *metabolomics* approaches that use liquid chromatography (LC) separation prior to MS analysis is a practical alternative to expand our current knowledge of LMW DOM in soil.[84, 85]

1.4   Mass spectrometry-based exometabolomics

Metabolomics aims to characterize all metabolites present in a biological system under a certain set of physiological conditions.[86] Recently, *exometabolomics* has emerged as an encouraging complement to metabolomics as it is aims to monitor the dynamic production and consumption of metabolites by characterizing the *extracellular* small

molecule environment.[87] The exometabolome can be measured over time, giving temporal data on compositional changes, or in the case of this work, across space as well. By treating the soil matrix as an extracellular biological system, exometabolomics can be used to characterize LMW DOM and its availability with, for example, depth or between sampling sites under a range of environmental conditions. Exometabolomics is thus a promising approach to provide a functional signature of soil microbial community activity, helping to identify hotspots of C vulnerability in Arctic systems.[88, 89] Although NMR,[90-92] and later GC/MS,[93-95] laid the foundation for metabolomics studies, liquid chromatography-mass spectrometry (LC/MS) has become a powerful approach for untargeted, global analyses of small molecules in complex biological systems for a variety of reasons.[96-99]

1.4.1 *Analytical figures of merit*

Since the unit of measurement is mass—more specifically, mass-to-charge ratio (*m/z*)—a universal, intrinsic parameter of comparison, mass spectrometry enables the analysis of organic molecules that vary in size, polarity, solubility, or thermal stability for example, all factors that complicate the effectiveness of alternative techniques. In addition, while there's not a single platform that can detect all LMW DOM species in soil at one time, LC/MS has recently enabled the detection and characterization of hundreds to thousands of organic compounds from soil in a single measurement, across a broad range of chemical classes (*i.e.* amino acids, sugars, nucleobases, lipids) and a wide mass range (50 – 2000 *m/z*), at nano- or even picomolar concentrations.[84, 85, 100, 101] Furthermore, the LMW DOM matrix is soil water, which makes LC an ideal separations platform as compared to GC, which requires volatilization and applies heat, or capillary

electrophoreses (CE), which is often less sensitive, has disproportionate responses to small variations in pH or temperature, and suffers from migration time variability.[102] A typical LC/MS-based metabolomics workflow is shown below in Figure 7.

Briefly, a liquid sample is introduced and separated on the LC column, where analytes are eluted, transferred into the gas phase (aerosol), and ionized by applying a voltage. Charged analytes are then directed into the mass spectrometer and focused into the mass analyzer (detector) by a series of lenses. Data output includes a chromatogram and mass spectrum, yielding two dimensions for annotation—a retention time (RT) and the *m/z* for the molecular ion (MS^1)—and quantitative information (relative intensity). After molecular ion detection, fragmentation or tandem mass spectrometry (MS^2) analyses may also be carried out, offering a third dimension for annotation or structural elucidation of unknowns.[103-106] There are many different types of LC columns and conditions (stationary and mobile phases), instrumentation and parameters (ionization sources and mass analyzers), each with various associated figures of merit (*i.e.* sensitivity, reproducibility, throughput). These figures of merit were assessed here, used to decide upon the platform employed in each study reported in this dissertation, and are described below in more detail.
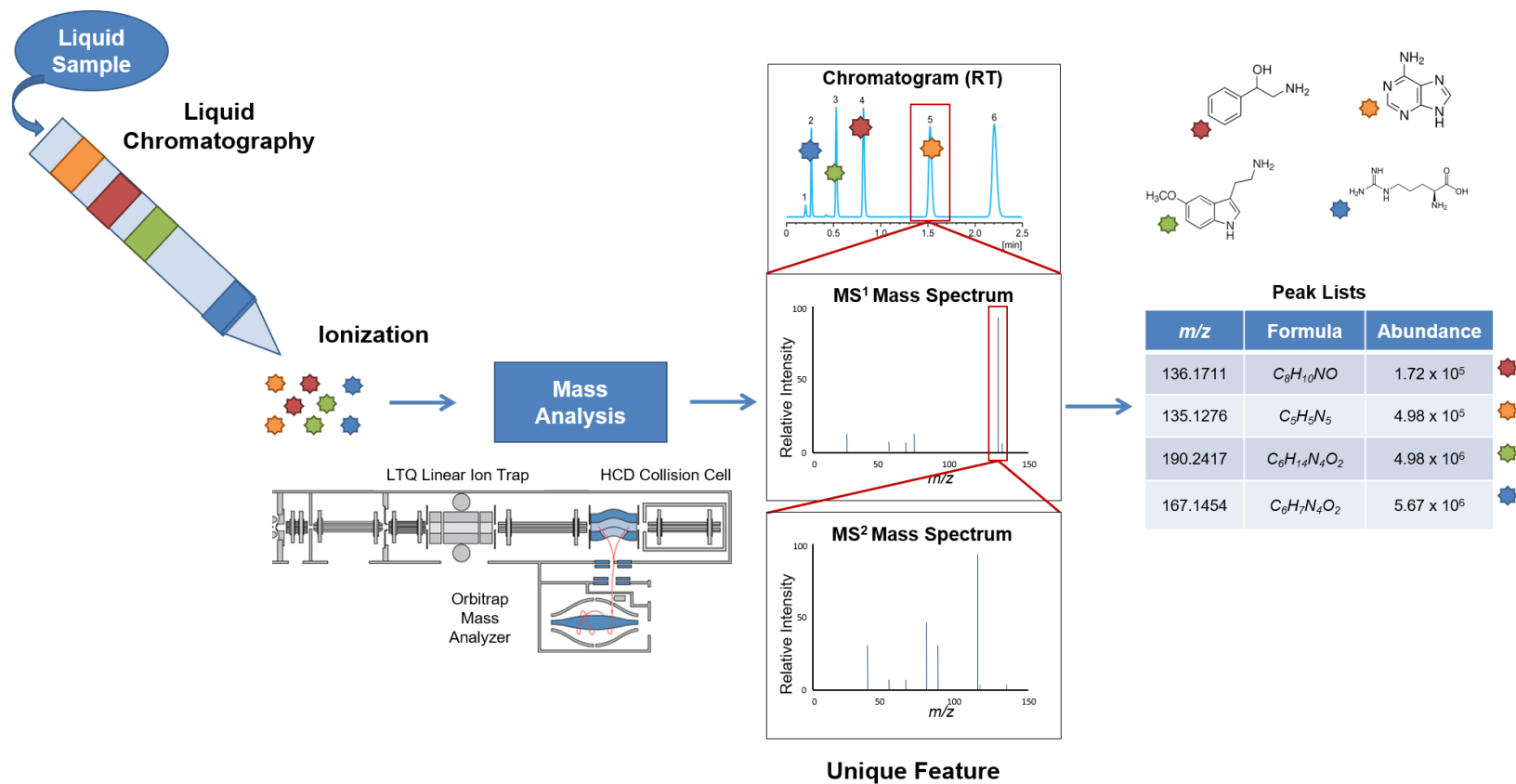
Figure 7: Typical LC/MS workflow from liquid sample introduction to molecular formula assignment

Source: LTQ-Orbitrap Velos Pro MS diagram obtained with permission from Thermo Fisher Scientific, Inc

The most common LC stationary phase used in metabolomics analyses is reversed-phase (RP) which employs a nonpolar, hydrophobic scaffold (*i.e.* C8, C18) to chemically adsorb hydrophobic compounds that are introduced in an aqueous mobile phase (Figure 8a). Analytes are eluted off the column by slowly increasing the concentration of a nonpolar, organic mobile phase. Because of this however, RP does not adequately retain small, polar molecules commonly found in biological mixtures, and in recent years, many new stationary phases have been introduced to combat this, each with varying selectivities; and some even include mixed-mode or multiple-interaction retention mechanisms.[92, 107]

For example, zwitterionic sulfoalkylbetaine phases—a type of hydrophilic interaction chromatography (HILIC)—contain both strongly acidic sulfonic acid groups and strongly basic quaternary ammonium groups bonded to a polymer backbone (Figure 8b). This enables multiple types of chemical interactions between various analytes and the stationary phase, increasing the number of compounds that can be retained.[108-110] These HILIC phases operate in reverse to the RP retention mechanism in that 1) analytes are introduced in an organic mobile phase, 2) polar, hydrophilic analytes are retained by a combination of electrostatic and hydrophilic interactions, and 3) are then eluted off the column by increasing the aqueous mobile phase conditions. Because RP and HILIC phases operate complimentary to one another, they are often paired together in metabolomics analyses, allowing the analyst to dig deeper into the metabolome and enhance overall coverage.[111-115]
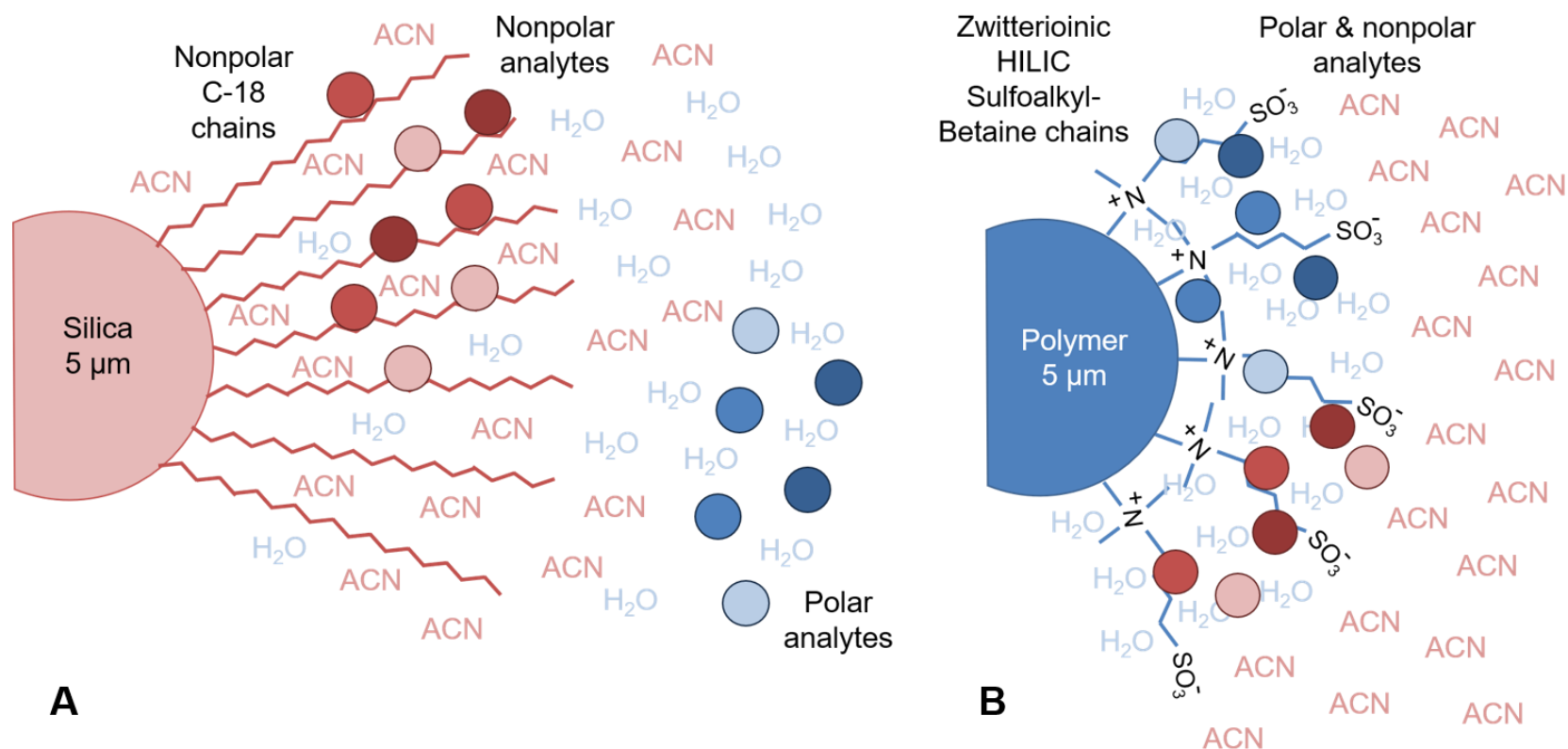
Figure 8: Schematic of (A) a C-18 RP silica bead and (B) a ZIC-pHILIC polymer bead showing the complimentary retention mechanisms where nonpolar analytes are absorbed to the C-18 chain and both polar and nonpolar analytes are partitioned into an aqueous layer formed on the surface of the zwitterionic chain.

While most MS-based environmental metabolomics analyses have been performed using GC/MS—possibly due to the lower cost for instrumentation and more widely available/more curated metabolite databases—LC has also been applied in soils extensively. Reversed-phase LC has dominated this space, largely, to characterize contaminants, pesticides, and other environmental contaminants.[116-119] However, HILIC columns have also recently been employed to detect LMW dissolved organic nitrogen standards recovered from a grassland soil,[84] amino acid standards from multiple temperate soils,[120] chitin-derived glucosamine to estimate fungal biomass in soil,[121] and LMW DOM in temperate soils to analyze mineral adsorption mechanisms[94] and define a soil media for microbial cultivation.[85] In addition, the combination of RP and HILIC has been applied in untargeted analyses of DOM from oceans, rivers, and streams (*i.e.* natural waters), which was recently reviewed by Sandron *et al.* 2015.[122] However, an untargeted dual-nano-LC/MS-based approach for the characterization of LMW DOM from Arctic soils has not yet been examined.

After LC separation, analytes are aerosolized and ionized; here, using electrospray ionization (ESI), which is performed at atmospheric pressure and can be directly coupled to LC platforms making it an ideal ionization source for LMW DOM measurements.[83, 123] Although first demonstrated in the 1970s,[124, 125] ESI didn't become commercially available until the 1990s, where it then helped transform many scientific disciplines, enhancing detection limits and expanding dynamic range (largest/smallest detectable signal).[126] As a "soft" ionization technique, ESI allows for the molecular ion to be detected by applying a high electric field (1-6 kV) and creating singly- or multiply-charged gas-phase ions,

effectively expanding the range of molecules that may be detected in a single measurement.[127] The ESI mechanism is further described in Chapter 2 below. Nano-ESI uses reduced LC flow rates and smaller dimensions (μm inner diameter columns vs. mm used in capillary columns). As such, it requires less sample (only nL vs. mL required by alternative techniques) and solvents, improves baseline separation,[128] and reduces the effects of ionization suppression from salts, improving sensitivity in detection over typical ESI by several orders of magnitude[129, 130] and making it an attractive approach for soil analyses where interference from inorganic salts is common.[84, 131]

Molecular ions are then directed into the mass spectrometer where they are separated based on their *m/z*, either in time or space. Mass analyzers have varying levels of resolution (ability to distinguish between different *m/z* ratios), sensitivity (signal-to-noise ratios), and data acquisition time (duty cycle). Generally, there is a trade-off between speed and resolution, because as the scan speed is slowed or the accumulation time maximized, mass accuracy (error between true *m/z* and measured *m/z*) improves. The advent of high-resolution mass analyzers like the Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap instruments has greatly improved the capabilities of MS platforms for differentiating complex mixtures of analytes, allowing for mass measurements out to four or sometimes five decimal places of accuracy and enabling putative elemental formula assignments to be made.[92, 132-134] While FT-ICR instruments provide the highest mass accuracy and resolution—sub-part per million (ppm) accuracy and 100,000 – 10,000,000 full-width at half-maximum (FWHM) resolution[135, 136]— Orbitrap instruments (Thermo Fisher Scientific) now routinely achieve < 5-ppm accuracy

and 1,000,000 FWHM resolution, are available as benchtop units, are more widely accessible, and more affordable.[137] As such, two Orbitrap instruments were used here and are further described in Chapter 2 and compared with the other main mass analyzers that dominate metabolomics research.

It is important to note that even with the most sensitive or the highest resolution instrumentation, *identification* of metabolites or absolute quantitation (*i.e.* targeted metabolomics) requires either isotopically-labeled standards or comparison to a matching authentic standard on the same system.[138] Given that authentic standards are frequently unavailable and identification requires a substantial investment of both time and resources, pooled sample quality controls (QC) and *annotation* by matching MS^1 data to online metabolite databases is routinely utilized for untargeted analyses seeking to distinguish biologically-relevant compounds first.[70, 137]

Another way to add confidence in formula assignments or database annotations, is with high-resolution tandem mass spectrometry data (MS/MS). After MS^1 mass analysis, additional structural information can be generated by isolating the molecular ion and fragmenting it, for example, by colliding it with an inert gas such as He or Ar—collision-induced dissociation (CID). This process activates or excites the molecules via multiple collisions and the kinetic energy generated is converted into internal vibrational energy within the molecule. At a specific energy threshold, the weakest molecular bonds break, or *fragment*, creating a chemical fingerprint unique to that molecule, the MS^2 spectrum. This is particularly useful for classifying unknown compounds—compounds that were not

assigned a molecular formula or did not match to a database—and identifying adducts (*i.e.* salts such as Na^+ or Cl^-) or complexes that formed during the electrospray process.

1.5   Dissertation overview

Comprehensively characterizing the soil biological system at the molecular level using an untargeted LC/MS-based approach allows for emergent ecosystem properties and processes to be discovered and defined. While technological advances in instrumentation have provided opportunities for improved chemical analyses in this space, the potential benefits of those technologies cannot fully be realized until it has been optimized and evaluated across a broad range of applications. In addition, optimizing a decided-upon LC/MS platform for the specific matrix in question is essential if reliable qualitative and quantitative information are to be obtained.[83, 137]

To that end, the principal goal of this dissertation was to address two primary research questions: 1) Can we sensitively and robustly detect and quantify LMW DOM chemistry across space in Arctic soils using untargeted LC/MS-based exometabolomics? and 2) What is the distribution of LMW DOM chemistry across a range of landscape features and conditions?

To address these fundamental research questions, here I have designed, implemented, and then evaluated an experimental workflow, from sample collection in the field, to data analysis and interpretation, and then applied the optimized approach across a range of Arctic landscape conditions and locations as part of the Next-Generation Ecosystem Experiments Arctic (NGEE-Arctic) project, a Department of Energy-led

initiative that aims to combine observational data and modeling approaches to fully integrate various established models, and ultimately, reduce uncertainty in climate model predictions. Site selection, experimental design, data processing and interpretation were all completed with this overarching aim in mind.

Briefly, samples were collected from two contrasting Alaskan field sites and with two approaches: an established technique in soil science known as a "destructive harvest," where soils are removed from the system and extracted with a liquid solvent, and an alternative, non-destructive technique that uses tension lysimetry to passively collect soil water *in situ* without disturbing the native soil environment. Liquid chromatography-mass spectrometry methods were optimized and applied to biological and/or analytical replicates from each site and collection method, resulting in thousands of LMW DOM features detected per sample. Multiple data analysis pipelines and software were evaluated for their capabilities to handle complex datasets, and user interfaces were also evaluated to ensure the techniques developed here could be widely accessible across a range of scientific disciplines. Datasets generated were investigated using multivariate statistical tools like analysis of variance (ANOVA), principal component analyses (PCA), and hierarchical clustering so show both qualitative and relative quantitative similarities or differences between samples or sites. Data obtained were analyzed to identify the LMW DOM features that were differentially-abundant between samples of varying depth, landscape topography, aboveground vegetation, or levels of permafrost degradation (natural thaw gradient).

25

This dissertation begins with a review of relevant literature covering mass spectrometry-based metabolomics methods to characterize small molecules in complex biological matrices (Chapters 1/2). Chapter 3 describes the development and evaluation of the untargeted, dual-LC (RP/HILIC), dual-polarity (positive- and negative- ionization mode), nano-ESI-MS/MS exometabolomics approach to characterize LMW DOM in Arctic soil and demonstrates the utility of the approach in detecting relative quantitative variations across space in soil (with depth). Chapters 4 and 5 report findings from applying the optimized technique across the Arctic landscape where the effects of topography, vegetation, and level of degradation (thaw) on LMW DOM availability are considered. In addition, a unique and important aspect of the Energy Science and Engineering doctoral program is the incorporation of an interdisciplinary focus. As such, a core aim of my graduate work was to contextualize the scientific research presented in this dissertation from a policy perspective. To accomplish this, in addition to taking courses in the areas of energy and environmental policy, I completed a 10-week internship in the Science & Technology Innovation Program with the Woodrow Wilson International Center for Scholars in Washington, D.C. As a result of this experience and follow-up research, Chapter 6 critically evaluates three contrasting U.S. policy alternatives for addressing the impacts of climate change in the Arctic, including an analysis of how *-omics* technologies can inform Arctic science and policy. Finally, Chapter 7 details the conclusions of these studies and summarizes recommendations for future research.

Ultimately, the contents of this dissertation demonstrate the optimization and application of analytical techniques that use available chromatographic materials,

instruments, and data analysis software to optimize and examine their utility in expanding current knowledge surrounding a complex and dynamic analyte pool in a unique and sensitive system with potentially significant feedbacks to climate change. This work is the first demonstration of this untargeted dual-LC, dual-polarity nano-ESI-MS/MS approach in Arctic soil; it brings new evidence to bear on our understanding of DOM in Arctic soils, and lays the analytical foundation for how to identify hotspots of biogeochemical activity in these soils going forward, providing an information-rich chemical profile which may be used to help reduce uncertainty in process-based model predictions of carbon and nitrogen cycling in the Arctic.

CHAPTER 2: MATERIALS, METHODS, INSTRUMENTATION, AND BIOINFORMATICS FOR LC/MS ANALYSES OF SMALL MOLECULES IN SOIL

Parts of this chapter have been adapted from the following manuscripts:

Ladd, M.P., Abraham, P., Giannone, R., Hettich R. Evaluation of an untargeted nano-liquid chromatography, dual-polarity, tandem mass spectrometry approach to expand coverage of low molecular weight dissolved organic matter in Arctic soil. *Scientific Reports (in review).*

Ladd, M.P., Reeves, D., Poudel, S., Iversen, C.M., Wullschleger, S.D. Hettich, R.L. Untargeted exometabolomics reveals biogeochemical hotspots with vegetation and polygon type in arctic tundra soils. *Environmental Science & Technology* (*in prep*).

ML's contributions included: literature review, experimental design, sample collection, sample preparation, data collection, data analysis, manuscript writing and editing.

2.1   Experimental considerations for exometabolomics in soil

The soil exometabolome is typically described as the sum of all the metabolites being produced, released, or consumed, thereby acting as a direct measure or snapshot-in-time of the net metabolic state of a complex soil microbial community.[88] Here, LMW DOM is used to describe the pool of analytes being characterized by the exometabolomics approach, to emphasize that not all of the small molecules being isolated (*i.e.* that are available for mineralization and release as a GHG) are of microbial origin. However, the experimental considerations associated with an exometabolomics study apply here as well. So, in addition to the figures of merit described above, each step of the workflow from sample collection and preparation to data analysis and interpretation was evaluated.

For example, because exometabolomics takes a data-driven approach, it is of the utmost importance that the technique be able to reproducibly and robustly differentiate signal from noise to avoid false positives. In addition, LMW DOM is often found in low concentrations and its composition can change on the order of hours or even minutes, so maintaining sample integrity from collection in the field to analysis in the lab and

differentiating quantitative variations with a conservative statistical approach was a top priority. Similarly, environmental analyses often demand multiple replicates to enable statistical comparisons, and as such, it was essential that the technique developed and evaluated here also be high-throughput and require minimal sample. Due to the complexity of the sample itself, it was also important to reduce sources of contamination or interference wherever possible. An additional area of consideration included reducing selective bias in recovery during analyte extraction or collection, and finally, instrumentation and software were also evaluated for their capability to provide and/or process both qualitative and quantitative information, about a broad range of analytes, with a user-friendly graphical user interface (GUI). This was done because although mass spectrometry data collection often demands a specialist, a supplementary aim of this dissertation was to ensure the data produced by this optimized approach is accessible to a broad range of scientists from multiple fields, including those beyond the mass spectrometry community—ecology, biogeochemistry, or hydrology for example.

2.2 Study sites

Soil samples and field observations were collected from two contrasting Arctic sites in Alaska (Figure 8); the study sites for the NGEE-Arctic project.[139] These Alaskan field sites were chosen based on their representativeness of common Arctic landscape types and whether they have certain environmental gradients that could be used as proxies to scale measurements between various models. The first site (Figure 9a), where samples were collected for Chapters 3 and 4, is on the North Slope of Alaska and was chosen to represent a cold, continuous-permafrost, polygonal tundra site. Near the village of Utqiaġvik, AK
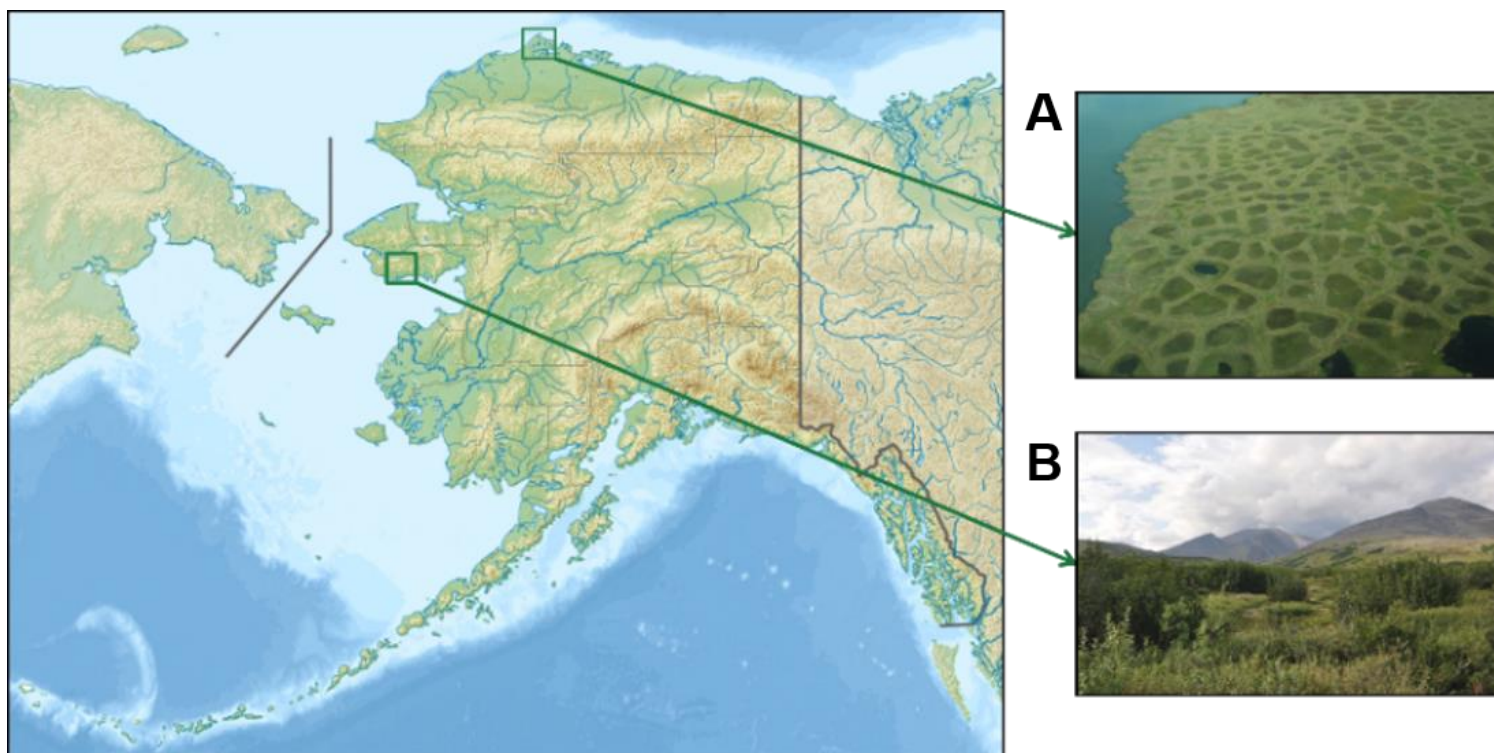
Figure 9: Map of Alaska showing two field sites selected by the NGEE-Arctic team, (A) a polygonal tundra site on the northern coastal plain near Utqiaġvik, AK and (B) a heterogenous, sub-Arctic terrain on the Seward peninsula inland from Nome, AK.

Source: Image courtesy of U.S. Department of Energy, Oak Ridge National Laboratory

(formerly Barrow), this site is dominated by characteristic landscape features like ice-wedge polygons (described in detail in Chapter 4) and drained thaw lake basins (DTLBs), that act as recognizable and quantifiable landscape units, that help to scale measurements and parameterize process models. The second site (Figure 9b), where samples for Chapter 5 were collected from, was established at a location south of the Arctic circle on the Seward peninsula, characterized by warm, discontinuous-permafrost and a more heterogeneous landscape with some polygons and DTLBs, but also well-defined watersheds and thaw (degradation) gradients, representative of future ecological and climate conditions under continued warming conditions. Preliminary measurements and observations taken by NGEE team members indicate this second site has more vulnerable carbon stocks, faster rates of vegetation change, and larger and more variable disturbance regimes (*i.e.* instances of fire or thaw/degradation).

2.3   Sample collection

In designing the experimental approach for sample collection, there were many aspects to be considered. Even during the summer months, the Arctic can be a very unpredictable, and at times, an unforgiving environment, with harsh conditions that make logistics and planning that much more challenging. Although there are flights that regularly fly through Utqiaġvik, it is still prohibitively expensive to ship supplies into or out of the field sites. As such, shipping was minimized, and all equipment and samples were transported as passenger luggage where possible. This required a detailed review of logistical challenges associated with maintaining sample integrity, reducing sources of contamination, ensuring efficient transport, and following all regulatory agency guidelines

and restrictions. As for the sample collection itself, based on a review of current literature, two techniques were chosen and have been described below. Use of blanks and controls, and replication strategy are discussed in each of the subsequent chapters.

2.3.1 Destructive harvests

The dominant collection technique for soil chemical analyses is the destructive harvest with subsequent solvent extraction. This technique requires that soil be removed from the native environment (*e.g.* soil cores, soil pits) and brought back to the lab for processing and analysis. Here, this approach was used in Chapters 3 and 4 to optimize the LC/MS technique, compare it to alternative approaches, and evaluate variations in LMW DOM variability with depth, polygon type, and aboveground vegetation. Soil cores (n = 4, 20-30 cm depth, 10 cm diam.) were obtained using a push-corer and a long knife. Mineral soil was visually identified and removed by hand in the field along with any loose vegetative material. Cores were immediately sealed in gallon freezer-bags (Ziplock), stored on blue ice for transport to a -20 °C freezer to slow microbial metabolic activity until field work was completed. Cores were then transported frozen from Alaska to Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee using blue ice and a sealed cooler and stored at -80 °C until processing.

2.3.2 Passive sampling

Given that the destructive harvest/extraction approach can significantly impact soil biogeochemistry,[140] and a core aim of this dissertation was to provide a high-throughput measurement of C vulnerability across Arctic landscapes, a second "nondestructive" or passive sampling approach was applied in Chapter 5. Here, we used tension lysimetry, with

mini-rhizons (Figure 10), which are small (1.5 mm i.d.), easy to install, and continuously and passively collect and partially-filter soil pore water by slowly (1 mL/min) passing it through a porous PVC tube into an evacuated container via negative pressure and capillary action. Soil pore-water collections were stored on blue ice in field, in a -20 °C until field work had finished, and in a -80 °C freezer back at ORNL until processing.

2.3.3   *Soil moisture, root weight, and soil C and N measurements*

Soil water content (Equation 4) measurements were made using a gravimetric soil moisture technique for Chapters 3 and 4 or using a 5TE soil moisture probe (Decagon Devices) in Chapter 5 for measurements in the field just prior to sample collection. Gravimetric analyses were completed by drying a subsample of soil (4 g) to constant weight in an oven at 105 °C for 48 hours.

$$\text{Equation 4: } \%\ soil\ moisture = \frac{Fresh\ wet\ weight - dry\ weight}{dry\ weight} * 100$$

To obtain gross estimates of live root biomass (root weight, g), live roots (determined visually by color and roundness/diameter) were removed and set aside during homogenization. Homogenization was limited to 20 min to reduce human-derived variation in the number of roots removed from each soil. Roots were dried to constant weight at 60 °C for 24 hours. Total organic carbon (TOC) and total nitrogen (TN) data were collected in triplicate on a Shimadzu TOC-L CSH/CSN analyzer (Columbia, MD). Briefly, a subsample of soil (2 g) or soil pore water (24 mL), is introduced to the instrument where it then transferred to a combustion tube. For TOC analyses, both pure and an acidified sample are analyzed to obtain a total carbon (TC) and an inorganic carbon (IC)
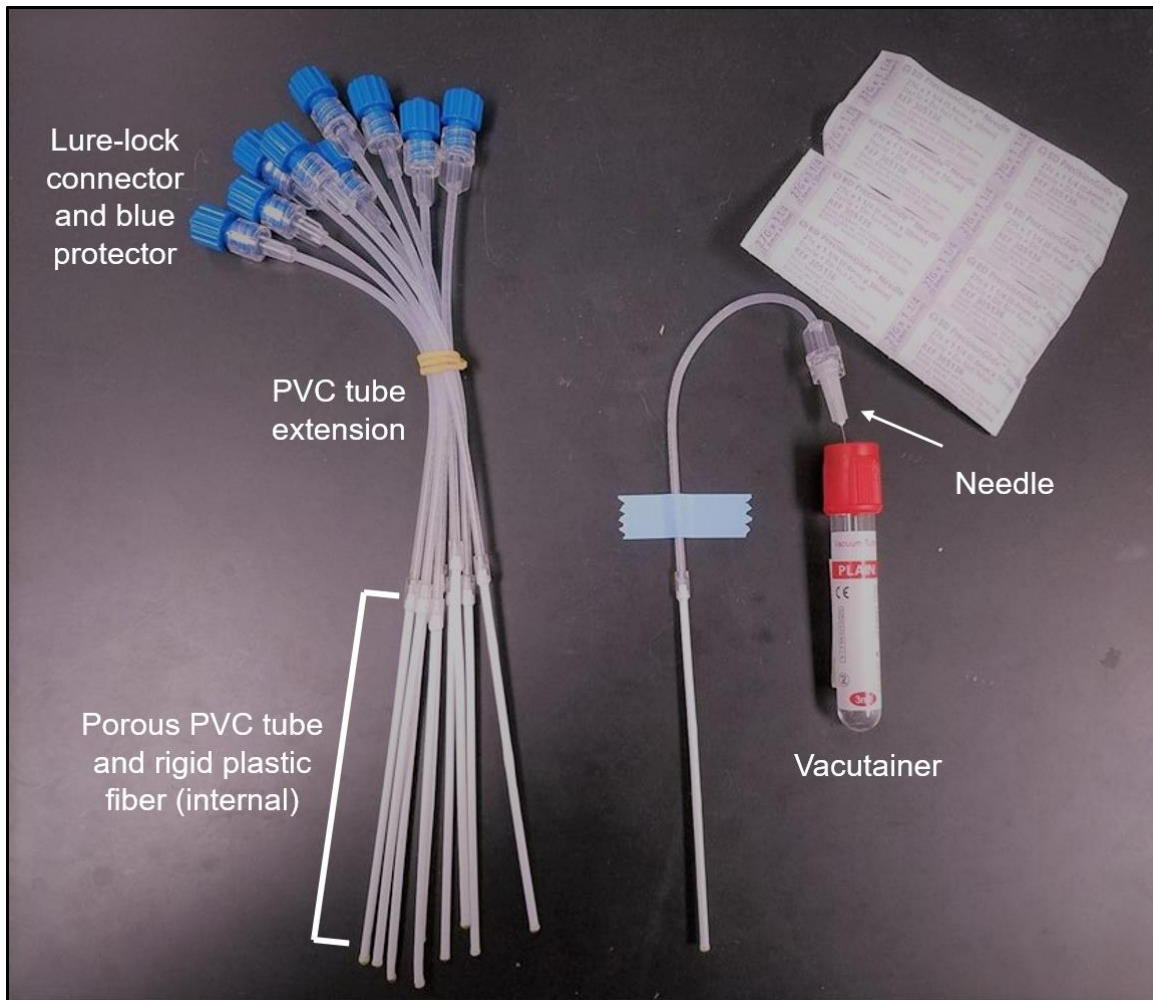
Figure 10: Photo of mini-rhizon samplers showing the porous PVC tube that is installed in the soil and the PVC extension that sticks out above ground, where a needle and vacutainer are attached to passively collect soil pore-water.

measurement, respectively, which can then be used to calculate TOC (TOC = TC – IC). A carrier gas (zero-carbon air) flows at 150 mL/min to the combustion tube, which has been filled with an oxidation catalyst (platinum) and is heated to 680 °C. The TC or IC of a sample is combusted into CO_2 which is then carried to a dehumidifier, where it is cooled, dehydrated, and detected using nondispersive infrared gas analysis (NDIR). The analog detection signal of the NDIR forms a peak which is proportional to the TC concentration

of the sample. Using a standard TC solution, a calibration curve is generated, and unknown TC concentrations may be calculated. For TN analysis, samples are introduced into the combustion tube packed with a catalyst (platinum) and the furnace temperature is set to 720 °C, creating nitrogen monoxide (NO) gas. Zero-carbon air is used to carry NO to the chemiluminescence analyzer where the NO reacts with ozone (O_3) creating products that are then measured photo-electrically generating a peak proportional to the total nitrogen concentration in the sample. Unknown concentrations are determined using a calibration curve as well.

2.4   LMW DOM extraction

2.4.1   *Choice of solvent, duration, temperature*

There have been many protocols developed for liquid extraction of DOM from soil, including but not limited to, aqueous or organic extractions, salt extraction with ammonium bicarbonate, KCl, or K_2SO_4 (0.5 – 2 M concentration), or a hot-water or methanol extraction.[140-142] The choice of solvent (including its pH), in addition to how long the extraction is carried out (using a shaker table), and at what temperature, have all been shown to impact the analyte pool that is ultimately extracted.[69, 75, 140, 141, 143] Here, in a preliminary analysis using Arctic soil, an aqueous extraction was compared with both a methanol extraction and two different salt extractions: KCl, as that is a common extractant used in soil DOM analyses, and ammonium bicarbonate as it is more amenable to electrospray ionization downstream. Soils were extracted for 1 and 24 hrs at room temperature and 4 °C. It was determined here that an aqueous extraction, at a pH equivalent to the native soil conditions, for a short time period (1 hr), at lower temperatures (4 °C),

most effectively 1) reduced noise while enhancing chromatogram complexity and analyte signal strength, 2) extracted the most-available small organic molecules (with minimal desorption of analytes from the mineral phase), 3) quenched further microbial processing of LMW DOM, and 4) reduced analyte degradation.[144] Different extractants can lead to different compositions of DOM molecules. Extraction by an organic, basic (NaOH), or salt solution for example would lead to different types and sizes of compounds being extracted. Here, aqueous extraction was used to select for the small (< 1000 Da) and soluble compounds most likely to be able to be used by plant and microbial communities directly.[134, 142] However, larger peptides and other substances such as dissolvable humics (400 – 2000 Da), may still be extracted during aqueous extractions[81] as well as detected by the LC/MS approaches used here (RP and HILIC mass range = 50 – 3000 Da, ESI-MS mass range = 50 – 2000 m/z). While the upper size limit for peptide transport systems across microbial membranes has been estimated to be ~600 Da,[29, 145] microbes can use extracellular enzymes to access a broader range of DOM substrates. As such, these larger, soluble compounds were not excluded from analysis as they are still considered available for microbial processing.

2.4.2   *Filtration and concentration*

To reduce possible sources of contamination and maintain high throughput, the direct analysis of both aqueous extracts and rhizon collections was also evaluated. Some sample preparation materials have coatings (*e.g.* polyethylene glycol, PEG) that ionize very well and "steal" charge from analytes-of-interest, obscuring the mass spectrum and effectively eliminating the reliability of any quantitative information that may have been

gleaned. So, fewer steps in the sample preparation workflow are ideal. However, even small soil particles left in a liquid sample could clog an LC column, especially at the nano-scale dimensions, and thus, sample filtration is often required prior to analysis. In addition, although a concentration step may introduce a selective bias for low volatility compounds, because LMW DOM is generally found in low abundances in soil, some concentration may be necessary to obtain detectable concentrations.

Here, we evaluated 3 kDa and 10 kDa filters from multiple manufacturers and visually inspected the amount of background signal after first use, and after a preliminary aqueous rinse. We also examined the effect of concentration-by-Savant (vacuum evaporation) using a mixed LMW DOM standard, a spike/recovery approach, and direct-infusion analysis.[146] We determined that the 3 kDa filter units from Amicon Ultra had an acceptable background after a preliminary aqueous rinse (neutral pH), and that while concentration (4 – 12x) was required to observe appreciable signal-to-noise (S/N) ratios for the destructive harvest/solvent extraction samples, rhizon collections could be analyzed directly. This may have been due to a dilution effect that occurred during liquid extractions—adding extra water on top of what was already present in the soil. Of course, there is always some compromise here, as concentration improves the signal of low-abundant analytes but can also over-enrich for a few dominant analytes. Since a primary objective of this work was to evaluate relative qualitative and quantitative differences between sites/samples, an optimized protocol was achieved for each study by observing the S/N ratios within each dataset and was maintained for the entire dataset to ensure consistency in the analysis (see subsequent Chapters' methods sections).

2.5   LC/MS analyses and instrumentation

In addition to selecting an appropriate stationary phase, column length, LC solvent and additives, and gradient conditions were also optimized and have been described here.

2.5.1   *Chemicals*

Acetonitrile (ACN), methanol (MeOH), isopropyl alcohol (IPA) and water (H_2O), all degassed and LC/MS-grade, were purchased from EMD Millipore (Billerica, MA, USA). Mobile phase additives including ammonium acetate (NH_4Ac), ammonium hydroxide (NH_4OH), and formic acid (FA), were purchased from Sigma-Aldrich. Authentic standards (> 98 % purity) representing a range of LMW organic compounds for analysis in Chapter 3 were purchased from Fluka-Honeywell Research Chemicals or Sigma-Aldrich. Stock solutions, 1 mmol L^{-1}, were dissolved in LC/MS-grade H_2O and standard curves were prepared by dilution with either ACN or H_2O, to match starting LC mobile phase conditions. Mixed standard solutions were prepared to final equimolar concentrations of 0.1, 1, 10, and 100 µmol L^{-1}. All stock solutions and dilutions were stored at -20 °C until analysis and FA or NH_4OH were added immediately prior to analysis.

2.5.2   *Liquid chromatography*

Measurements of standards and samples were carried out using a Dionex UltiMate 3000 HPLC pump (Thermo Fisher Scientific) coupled to either an LTQ-Orbitrap Velos Pro mass spectrometer in Chapters 3 and 4, or a Q-Exactive Plus mass spectrometer in Chapter 5 (both Thermo Fisher Scientific), each equipped with a nano-electrospray ionization source (Proxeon, Denmark) operated in positive- or negative-ion mode under direct control

of the XCalibur software (Thermo Fisher Scientific). The differences between these two instruments are described in detail in the following sections.

In Chapter 3, extracts were thawed and prepared immediately prior to injection by adding either 0.1 % FA or NH_4OH to help with ionization, and either 6-methylaminopurine riboside (6-MAP) or adenosine (final concentration, 10 µmol L^{-1}) as an internal standard for positive- or negative-ion mode, respectively. Internal standards were added to monitor method performance and reproducibility, and to assist with retention time alignment, relative quantitation, and annotation of LMW DOM.[147] While an internal standard for each feature detected would be necessary for absolute quantitation, only a single standard for each ionization mode was necessary here to evaluate the effectiveness of the technique at detecting relative quantitative variations across space (with depth) in a single core.[148, 149]

In Chapters 4 and 5, instead of a single internal standard, a pooled quality-control (QC) sample, consisting of equal volumes of all samples plus an internal standard, was prepared to monitor instrument performance and assist with normalization procedures used to evaluate and remove experimentally-derived variation between soil cores and sampling sites.[150] All analyses were randomized to minimize instrument-derived variation, and technical blanks representing the column re-equilibration conditions were run regularly to monitor background ions and carry-over. Controls (water extraction without soil and pure water collection through rhizon into vacutainer) were also analyzed in each study and used to subtract background and artifacts during data analysis (*i.e.* features that were from the sample preparation or analysis procedures and not analytes from the sample).

Separations were performed on 100 µm i.d. fused-silica (Polymicro Technologies) columns, which were laser-pulled in-house and pressure-packed to 20 cm with either Kinetex C18 resin (5 µm, 100 Å, Phenomenex) or zwitterionic, polymer-based ZIC-pHILIC resin (5 µm, Sequant, bulk material kindly provided by EMD Millipore) resulting in four separate LC/MS analyses per sample (HILIC +/- and RP +/-). Mobile phase compositions, gradient conditions, and MS parameters were systematically adjusted to provide the best ESI spray stability, signal strength, LC peak shape, and separation. Only mobile phase additives that were compatible with the ESI source were examined (Table 1). Thus, ion-pairing agents and non-volatile buffers were excluded from method development. The final gradients used for each LC/MS condition are listed in Table 2. Prior to MS analysis, each column was washed off-line for 1 h with an alternating gradient from 100 % A to 100 % B to expand the range of compounds that would be retained, but never exceeding a total composition of 60 % aqueous on the HILIC columns so as not to disrupt the aqueous layer on the surface of the stationary phase.[109, 151]

In Chapters 3 and 4, samples, standards, and QCs were manually injected directly onto the column using a 1 µL fused-silica loop, and in Chapter 5, an autosampler (Ultimate 3000 RS, Thermo Fisher Scientific) was used. Nano-flow rates were achieved with a split-flow setup prior to injection (20 nL). The pump was set to 0.150 mL min^{-1}, measuring ~250 nL min^{-1} at the tip. A post-gradient wash was applied at the end of each run to ensure column re-equilibration and maintain the ionic strength of the HILIC material.

Table 1: Mobile phase conditions and additives that were tested to optimize each LC phase and MS polarity. Final mobile phase compositions are shown in bold font.

HILIC (+)		
A: 60 % ACN, 40 % NH₄Ac, 0.1 % FA	**B: 95 % ACN, 5 % NH₄Ac, 0.1 % FA**	pH
2.5 mM NH₄Ac	2.5 mM NH₄Ac	3.2
5 mM NH₄Ac	**5 mM NH₄Ac**	**3.5**
10 mM NH₄Ac	10 mM NH₄Ac	3.7
20 mM NH₄Ac	20 mM NH₄Ac	4.0
HILIC (-)		
A: 100 % NH₄Ac, 0.1 % NH₄OH	**B: 95 % ACN, 5 % NH₄Ac, 0.1 % NH₄OH**	pH
2.5 mM NH₄Ac		9.0
5 mM NH₄Ac	**5 mM NH₄Ac**	**9.1**
10 mM NH₄Ac		9.2
20 mM NH₄Ac		9.3
RP (+)		
A: 95 % H₂O, 5 % ACN	**B: 70 % ACN, 30 % H₂O**	pH
0.1 % FA	**0.1 % FA**	**3.5**
RP (-)		
A: 97 % H₂O, 3 % MeOH	B: 100 % MeOH	pH
20 µM TEAB*	20 µM TEAB*	8.3
15 mM acetic acid		5.0
A: 90 % H₂O, 10 % IPA	**B: 80 % ACN, 10 % H₂O, 10 % IPA**	pH
1 mM NH₄OH	**1 mM NH₄OH**	**9.0**

*tetraethylammonium bromide

Table 2: Optimized gradient conditions for nano-LC separations, for positive- and negative-MS ionization modes on C18-RP and ZIC-pHILIC columns

C18 Reversed-Phase				ZIC-pHILIC			
Positive		Negative		Positive		Negative	
time, min	% B	time, min	% B	time, min	% A	time, min	% A
0.0	2	0.0	25	0.0	0	0.0	0
3.0	2	3.0	25	3.0	0	3.0	0
23.0	100	23.0	100	23.0	100	23.0	30
28.0	100	28.0	100	28.0	100	28.0	30
33.0	2	33.0	25	30.0	80	30.0	60
40.0	2	40.0	25	35.0	80	35.0	60
				40.0	0	40.0	0
				45.0	0	45.0	0

2.5.3   *Nano-electrospray ionization*

Each column was then positioned on the nano-spray source aligned in front of the MS inlet (Figure 11). A voltage is applied directly prior to the column, so as solvent droplets leave the tip of the column, they quickly dry, creating an aerosol of tiny charged droplets that propagate out forming a *Taylor cone* of even smaller droplets (Figure 12). As the solvent evaporates, analyte ions form when the charged droplets reach their *Rayleigh limit*—when the electrostatic repulsion becomes more powerful than the surface tension of the droplet—where they then undergo a *Coulomb explosion* forming tinier and tinier droplets (Figure 12). A *nebulizing gas* (*i.e.* nitrogen) can be used to assist with drying, and the heated capillary (~250 °C) also removes trace amounts of solvent remaining as the charged analytes enter the mass spectrometer.
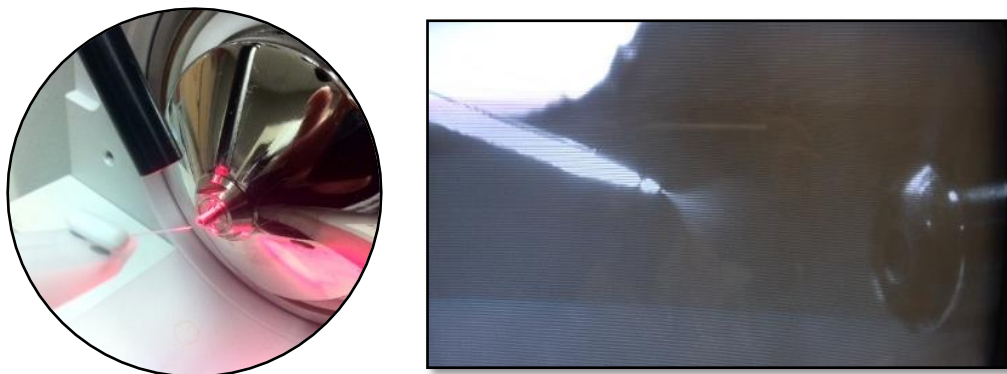
Figure 11: Optical photo of nano-spray setup with column aligned in front of heated MS capillary inlet (left) and magnified capture of the electrospray Taylor cone being formed in front of the inlet (right).
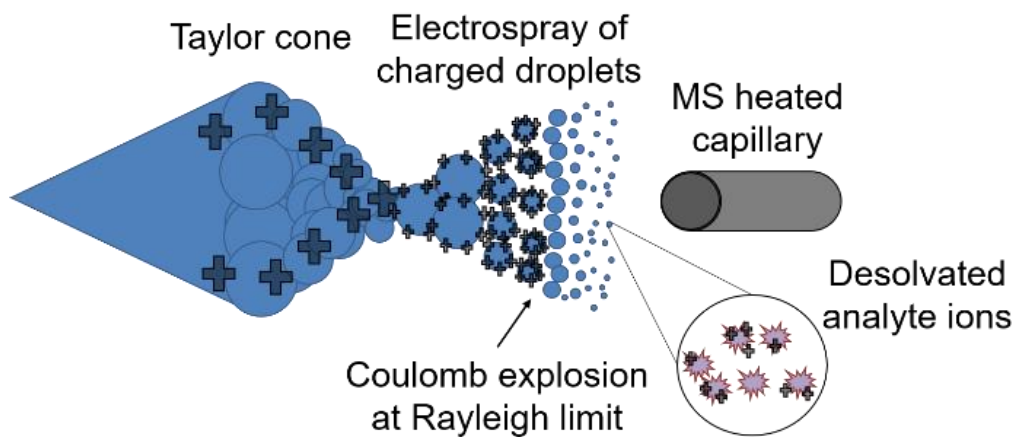
Figure 12: Schematic of electrospray mechanism in positive-ion mode showing droplet drying, aerosol formation, Coulomb explosion, and charged-ion formation

2.5.4   *Mass analysis*

After ionization, positively- or negatively-charged ions are focused into the mass analyzer using a series of lenses with successively increasing voltages. In metabolomics, there are four predominant mass analyzers that are used:

1) quadrupoles that use electric or magnetic fields to scan across a user-defined mass range detecting ions of increasing *m/z* as they move through space (*e.g.* linear quadrupole, triple quadrupole, QQQ),

2) ion trapping analyzers where ions are accumulated in a two-dimensional cell by a radio frequency (RF) and direct current (DC) applied to the front and back of the trap electrodes; mass analysis occurs sequentially in time by increasing the RF voltage and systematically ejecting ions from the electrostatic trap when they become unstable (*e.g.* linear ion trap, quadrupole ion trap, LTQ),

3) time-of-flight (TOF) mass analyzers detect ions by their differing flight times, rather than scanning across a mass range, separating them by either their temporal (*i.e.* two ions of the same mass are formed at different times and arrive at the detector at different times) or spatial (*i.e.* two ions of the same mass are formed in different locations and arrive at the detector at different times) distribution, and

4) orbital frequency mass analyzers that detect ions oscillating within a cell due to an applied electric or magnetic field, where each mass assumes a unique frequency of rotation directly related to its mass-to-charge ratio (*e.g.* FT-ICR, and Orbitrap).

There are also hybrid instruments that combine mass analyzers (like the Orbitrap Velos Pro used in Chapters 3 and 4), that allow for tandem mass spectrometry experiments (MS/MS) to be carried out (either CID or higher-energy collisional dissociation, HCD), where a molecular ion is isolated in one detector using data-dependent acquisition (DDA) and then fragmented and detected by another (tandem-in-space). This allows for separate resolutions (low vs high) to be set for the MS^1 and MS^2 measurement, enabling a more optimal duty cycle,[133, 137] however this was not applied here. Both MS^1 and MS^2 measurements were completed using CID and high-resolution detection in the Orbitrap. The most common type of DDA is the "TopN" mode where, for example, the top five most abundant ions are sequentially isolated and fragmented, and then placed on an exclusion list for an indicated set of time so they are not resampled allowing for a deeper measurement. The parameters used for fragmentation in each study have been listed in each of the subsequent chapters.

In deciding on the detector to use here, each of the figures of merit described in Chapter 1 (*i.e.* mass accuracy, mass resolving power, dynamic range, sensitivity, and duty cycle) were examined and optimized for the analyte- and matrix-of-interest. For Chapters 3 and 4, a hybrid LTQ-Orbitrap Velos Pro was used, and then, for Chapter 5, work transitioned to a Q-Exactive Plus, which was chosen for its faster scan speed and higher resolving power (Table 3). This allowed for more features to be differentiated and ultimately annotated, which are attractive characteristics for metabolomics analyses of complex biological matrices. Diagrams of the LTQ-Orbitrap and Q-Exactive Plus instrumentation are shown in Figure 7 (above) and Figure 13 (below), respectively.

46

Table 3: Figures of merit between the two MS instruments used in this work

Figures of Merit	Orbitrap Velos-Pro	Q-Exactive Orbitrap
Resolving power	60,000 (FWHM) at 400 *m/z*	140,000 (FWHM) at 200 *m/z*
Scan speed	1 scan/sec	12 scans/sec
Dynamic range	>5000 between highest and lowest detectable ion	>5000 between highest and lowest detectable ion
Mass range	50 – 2000 *m/z*	50 – 6000 *m/z*
Mass accuracy	<1 ppm with internal calibration	<1 ppm with internal calibration
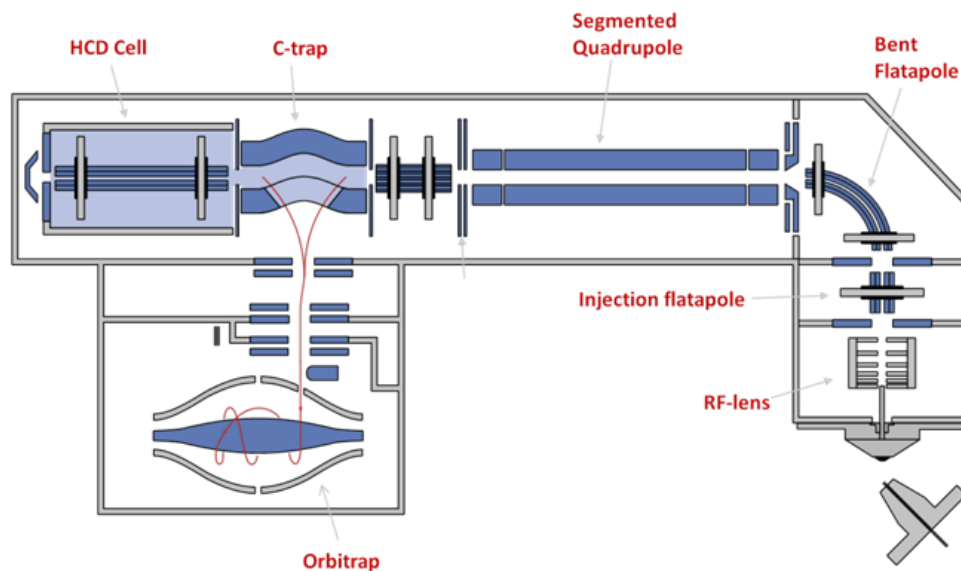
Figure 13: Schematic of Q-Exactive Plus mass spectrometer

Source: Obtained with permission from Thermo Fisher Scientific, Inc.

2.6 Data extraction and processing

Due to the high complexity of LMW DOM, the multiple LC/MS conditions used, and the fact that high-resolution measurements can differentiate molecules that vary in mass by less than one mass unit, this untargeted approach produces very large datasets, frequently with thousands of peaks detected in a single sample. As such, developing a conservative data filtering and analysis approach was integral to ensuring accurate interpretation. Here, I describe our optimized approach for filtering out baseline noise and false-positives, and for identifying LMW DOM features that were consistently, and significantly differentially-abundant between samples and/or conditions.

2.6.1 Peak detection and alignment

Raw LC/MS data were subjected to peak picking, alignment, and normalization using MZmine2 (v2.30).[152] This software is open-source and has a user-friendly GUI with separate modules for each data processing step, but also includes a batch-processing mode, maximizing the accessibility of the software's capabilities to new users or experienced analysts alike. A detailed description of each of the modules used for data analysis is listed in Appendix A, and the optimized module parameters and data filtering strategy established here were as follows:

Prior to statistical analyses, it is important in untargeted analyses to be able to detect as many small, but real analyte signals as possible. Here, differentiating between true and false signals was accomplished by first optimizing three parameters in the MZmine peak extraction algorithm—minimum peak height, MS^1 tolerance, and RT window. These parameters, which have been reported for each study in the subsequent chapters, were

optimized by manually inspecting the accuracy of peak assignment. Precursor ions that were selected for fragmentation were identified with the MS/MS peak list builder (+/- 0.005 *m/z* or 10 ppm MS^1 window) followed by the peak extender module which searches for data points in both directions of the RT apex (MS^1 mass tolerance +/- 10 ppm, intensity > 1.0E5). Isotopic peaks (*i.e.* ^{13}C natural abundance ion, mass difference of a neutron = 1.0033 Da) were then removed with the isotopic peaks grouper module using a +/- 0.001 *m/z* and 1 min RT tolerance in order to avoid errors with relative quantitation and annotation. During the ESI process, while less likely than with other ionization techniques, in-source fragmentation can occur, along with the formation of non-proton adducts with Na$^+$, K$^+$, or NH$_4^+$ for example, or complexes that coelute with analytes of interest. Here, fragments were identified in MZmine by comparing peak lists with MS^2 scan data (same *m/z* within +/- 5 ppm and same RT +/- 0.1 min), while adducts were identified in MZmine by the mass difference between the original ion and the adduct being equal to the mass selected by the user (*i.e.* +/- 5 ppm from 22.9892 *m/z* for a Na$^+$ adduct) and having a matching RT (+/- 0.1 min). Finally, complexes were identified in MZmine by searching for peaks with the same RT time (+/- 0.1 min) that add together to make the ion complex *m/z* (+/- 5 ppm). To help reduce any chromatogram shifts that would impact annotation, but include features whose RTs had shifted slightly between extraction replicates, peaks from the same chromatographic phase and ionization mode were aligned (+/- 5 ppm, +/- 2 min RT) based on 10 iterations and at least a 25 % match score using the nonlinear, random sample consensus (RANSAC) algorithm.[153, 154] Aligned peak lists were exported to .csv files for data filtering procedures.

2.6.2 *Normalization, data filtering, and relative quantitation*

To evaluate the ability of the LC/MS approach to detect quantitative variations in LMW DOM availability across space, in addition to peak detection and alignment, it is also important to remove as much noise, background signal, and unwanted variation as possible. To accomplish this, multiple conservative LC/MS-based metabolomic data processing techniques were applied, including normalization procedures, a blank/control correction, and reproducibility and abundance thresholds.[96, 98, 99] While there are many different methods for normalizing metabolomics data, each comes with various drawbacks and tradeoffs (*i.e.* bias-variance trade-off) and no single approach perfectly describes all the unwanted variation associated with an experiment, which is why it is important to consider the experimental design and aims of the study when optimizing a normalization approach.[155] For example, while normalizing to an internal standard that is specific to each compound-of-interest (targeted analyses) or to a mixed internal standard with compounds from multiple classes for untargeted analyses are alternative normalization approaches commonly used in metabolomics analyses, these require the introduction of several external compounds to the sample, which not only further complicate the chromatogram and mass spectrum, but could also alter the composition of the sample via chemical reactions. Here, integrated LC peak areas were obtained from the aligned extracted ion chromatograms (XICs), normalized to per gram dry soil (in Chapters 3 and 4) to account for moisture variations between samples, and then $\log_2$-transformed for ease of data interpretation. To control for systematic variation between samples and remove intragroup batch effects (Figure 14), the $\log_2$-transformed peak areas were also normalized to 1) an internal standard specific to each ionization mode in Chapter 3, using a ratio factor deter-
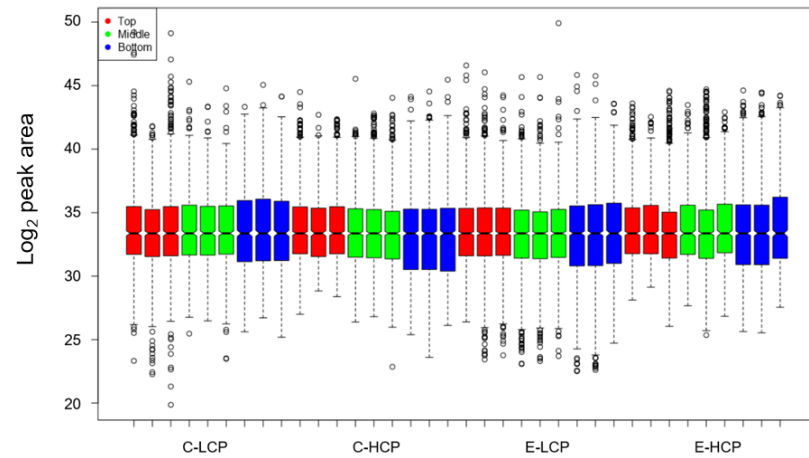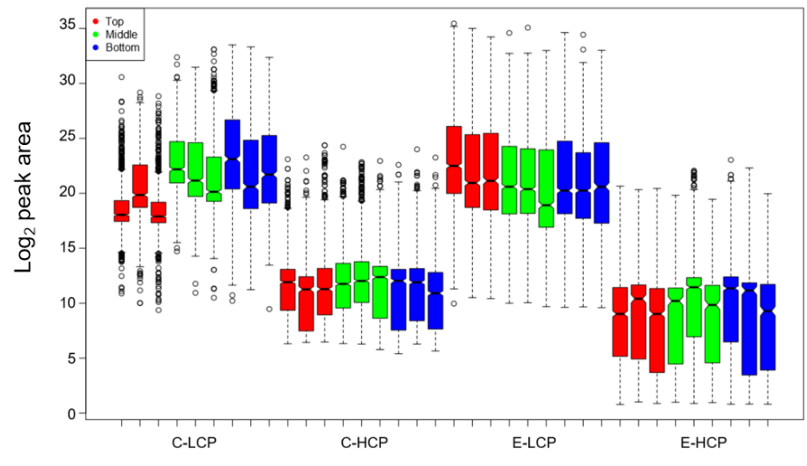
Figure 14: Box-and-whisker plots of (left) raw $\log_2$ peak areas for an example dataset obtained from the study described in Chapter 4 which shows a systematic shift in values between soil cores analyzed on different days and (right) the normalized $\log_2$ abundances showing the removal of experimental variation by normalization procedures (Chapter 2)

-mined with the MZmine standard compound normalizer module,[154] and 2) to pooled-sample QCs in Chapters 4 and 5 using QC-RLSC (robust LOESS signal correction),[156] with two scaling factor techniques, LOESS (locally estimated scatterplot smoothing) and median-centering, all completed in the freely-available InfernoRDN and R environments.[157]

By including controls and daily technical blanks, any artifact signals that originated from sample collection, preparation, or analysis (*i.e.* extraction leachates, solvent contaminants, column background) and were above a specified noise level could then be easily identified and manually removed, decreasing the false discovery rate (FDR) of the technique.[98, 99] This resulted in a matrix of *features*—defined here as a unique RT, MS^1 *m/z*, and MS^2 fragmentation spectrum with a corresponding peak height (intensity) and a peak area. Any duplicate features (same MS^1 *m/z* and peak area, but a different retention time due to alignment error) or features that had zero peak area after normalization were also removed, resulting in a matrix of *high-quality features* (HQFs). The number and complexity of HQFs detected by each LC/MS condition were used to evaluate LMW DOM coverage, measurement depth, and the qualitative and quantitative reproducibility across samples by comparing the accurate mass of the corresponding $[M+H]^+$ or $[M-H]^-$ molecular ion and the peak area for each feature. Next, only the HQFs that were observed in at least two of three replicates for the study described in Chapter 3, or three of nine for the studies described in Chapters 4 and 5, were carried on to subsequent quantitative analyses. This step helps reduce the probability of false positives and creates a more conservative list of only the most reproducible and *abundant HQFs* to be compared between samples. Missing

values were then imputed for statistical analyses by randomly selecting numbers from a normal distribution near the limit of detection (width = 0.3, downshift = 1.8-2.3) using the freely-available Perseus software.[158] Finally, various univariate (Student's t-test) and multivariate statistical analyses (ANOVA), and data visualization techniques (PCA, volcano plots, and/or heat maps) were used to help identify clusters of features that were consistently and significantly varying across the sample sets for annotation and to examine the relative abundance differences between extraction replicates and core depths (Chapter 3), polygon or vegetation types (Chapter 4), or along a natural thaw gradient (Chapter 5).

2.6.3   Statistical analyses

Across all the studies, variation between extracts or rhizon collections to assess reproducibility was analyzed using Pearson's correlations that were performed with JMP Pro (v13.1).[159] In Chapters 3 and 4, overall variation across the dataset was first visualized using PCA, which is an unsupervised, data dimension-reduction technique that plots the weighted-sum of the contribution of a set of LMW DOM features within a sample to a principal component and compares that to all the other samples. While PCA can be used as a multivariate statistical analysis, it suffers from the multi-colinearity problem that is common with metabolomics datasets, in that they generally have more dependent variables (*i.e.* metabolites, in the hundreds or thousands) than independent variables (*i.e.* biological conditions, in the tens). An alternative statistical approach that is frequently applied in metabolomic datasets is that of partial least squares discriminant analysis (PLS-DA) which alleviates the independent-to-dependent ratio issue.[160] However, PLS-DA is a *supervised* technique, in that it plots the variation in the dataset after first considering the correlation

between the dependent and independent variables. Thus, PCA was used here to first visualize the overall variation across the untargeted datasets.[161] Then, to determine differentially-abundant LMW DOM features, 1) Student's t-test was used to compare profiles between the three depths in Chapter 3 and between cores of the same polygon type or vegetation in Chapter 4, and 2) ANOVA—a multivariate statistical technique that analyzes the differences between groups using the means across replicates—was used to compare cores in Chapter 4 and thaw conditions in Chapter 5 using the Python SciPy library.[162] Tukey's range test was used as a post-hoc analysis to compare all possible pairs and identify abundance differences greater than the expected standard error between groups. Because pairwise comparisons by t-test lead to a multiple-testing error with metabolomic datasets, *volcano plots*—which consider the fold change (FC) between two conditions—were used to identify significant features that passed both a p-value threshold and a FC threshold.

For both the ANOVA and t-tests, any feature with a $\log_2$ fold change > 2 and a p-value < 0.05 was considered significant, but we also explored tighter parameters (*i.e.* $\log_2$ FC > 4, p-value < 0.001) to highlight LMW DOM features that were highly significant. In addition, two-way hierarchical-clustering (heat maps) using the Ward agglomerative technique were used to visualize these variations and select clusters of features that varied similarly across the dataset for annotation. Volcano plots and heatmaps were generated in Perseus, and PCAs were produced in the InfernoRDN environment.

2.6.4   *Annotation*

Annotation of features that were consistently observed and significantly, differentially-abundant due to depth, polygon type, vegetation, or thaw was carried out in a three-step procedure. First, features ($[M+H]^+$ or $[M-H]^-$ ions) were searched against multiple freely-available online using high mass accuracy measurements (precursor mass tolerance of 5 ppm) within MZmine and using the MetaboSearch tool.[163] Databases included KEGG,[164] METLIN,[165] MMCD,[166] PubChem,[167] HMDB,[168] LipidMaps,[169] or Plant Cyc.[170] While it depends on the database size, this first filter is the most powerful and generally can remove up to 99.9 % of false candidates.[171] Second, in chapters 4 and 5, putative chemical formulas were assigned using the MZmine elemental formula assignment module and the following criteria established using Kind and Fiehn's "Seven Golden Rules" and parameters modified from Kujawinski and Behn's compound identification algorithm (CIA) for small molecules:[172-174] mass measurement error of < 5 ppm, taking into account the presence of C_{1-100}, H_{3-100}, N_{0-30}, O_{1-50}, P_{0-3}, S_{0-3}, and elemental ratio heuristics including $0.1 <= H/C <= 6$, $N/C <= 4$, $O/C <= 3$, $P/C <= 2$, and $S/C <= 3$. When multiple candidate formulas were returned, to ensure that an objective choice was made, we consistently chose the formula with the lowest error, lowest number of heteroatoms, and if there was a phosphorus present, at least three oxygen atoms must have also been present in the formula.[175] Third, compounds that matched to multiple hits in a database were manually scrutinized in an iterative approach by assessing high-resolution mass spectral data for consistent fragmentation profiles, or by using the similarity matching tool in MZmine, to filter out false candidates and annotate unknown (unmatched) features.

It is important to note here that while we included an annotation step in these analyses, it would be outside the scope of this study to *identify* the LMW DOM features by matching to authentic standards as that would limit our analytical window to only metabolites that have been synthesized. Due to the complexity of this analyte pool, most of the features detected are likely "unknowns", and authentic standards are frequently unavailable. For the aim of distinguishing a profile of features (known or unknown) that were differentially-abundant across space, with the ultimate goal of linking that chemical profile to biological processes (*i.e.* methanogenesis) or as an indicator of C vulnerability, high-mass accuracy MS^1 and MS^2 annotations and putative identifications by database matching or elemental formula assignment were sufficient.

CHAPTER 3: OPTIMIZATION AND EVALUATION OF AN UNTARGETED EXOMETABOLOMICS APPROACH TO EXPAND COVERAGE OF LOW MOLECULAR WEIGHT DISSOLVED ORGANIC MATTER IN ARCTIC SOIL

The chapter presented below has been adapted from the following manuscript:

Ladd, M.P., Abraham, P., Giannone, R., Hettich R. Evaluation of an untargeted nano-liquid chromatography, dual-polarity, tandem mass spectrometry approach to expand coverage of low molecular weight dissolved organic matter in Arctic soil. *Scientific Reports (in review).*

ML's contributions included: literature review, experimental design, sample collection, sample preparation, data collection, data analysis, manuscript writing and editing.

3.1   Abstract

Characterizing LMW DOM in soils and evaluating the availability of this labile nutrient pool is critical to understanding the underlying mechanisms that control carbon storage and release across many terrestrial systems. However, due to wide-ranging physicochemical diversity, characterizing this complex mixture of small molecules and how it varies across space remains an analytical challenge. In this chapter, we optimized and evaluated an untargeted exometabolomics approach to detect qualitative and relative-quantitative variations in LMW DOM availability with depth using a soil core obtained from the Alaskan Arctic. We combined RP and HILIC liquid chromatography, and ESI coupled with high-resolution mass spectrometry (HRMS) in positive- and negative-ionization mode. Using a data-dependent approach, tandem mass spectrometry (MS/MS) experiments were also carried out, adding a third dimension (RT, MS^1, and MS^2) for annotation and flexibility in the technique to examine both known (already listed in a database) and unknown compound structures. Because soils have high salt concentrations which result in substantial ion suppression at the macro-scale, we employed a nano-scale LC column/emitter and flow rates to enhance sensitivity and enable more accurate relative quantitation. Establishing this methodology for the first time in Arctic soils lays the

technical foundation for future studies aiming to incorporate LMW DOM molecular data into mechanistic models.

3.2   Introduction

In recent years, LC-ESI-MS has become a powerful analytical tool for obtaining broad coverage of chemically-complex mixtures of small molecules in metabolomic analyses (see Chapter 1).[92, 176] While RP liquid chromatography in positive MS-ionization mode has dominated untargeted metabolomic studies, the limitations of using a single chromatographic phase or polarity have also been documented;[96] especially when analyzing mixtures with a high fraction of water-soluble, highly-polar metabolites,[112, 177] as these compounds are not well-retained by RP.[178] Hydrophilic interaction liquid chromatography (HILIC) however, has been shown to be an effective tool for retaining and separating small, highly-polar compounds, thereby enabling quantitation.[109, 179] In addition to combining multiple LC techniques, adding negative-ionization has also been shown to expand metabolome coverage in bacterial cultures, plant and human tissue, and urine.[97, 114, 180, 181] However, a dual-LC, dual-polarity untargeted exometabolomics approach has not yet been examined for the characterization of LMW DOM in Arctic soils. As such, in this study, we optimized and evaluated RP- and HILIC-ESI-MS in positive- and negative-ion modes for the characterization of LMW DOM from soil water extracts, and then applied the optimized technique along the length of an Arctic organic horizon to examine the capabilities of the approach in determining relative abundance differences across space (with depth).

3.3   Experimental approach

3.3.1   *Sample collection and processing*

A soil core (10 cm diameter, ~ 30 cm depth) was collected from the organic-rich active layer of a continuous-permafrost landscape, from the center of a low-centered polygon (see Chapter 4) on the Barrow Environmental Observatory (BEO), AK (71° N, 156° W) and shipped frozen to Oak Ridge National Laboratory (ORNL, Oak Ridge, TN) where it was stored at -80 °C until processing. The frozen core, representing a single organic horizon identified by visual inspection of the soil layers, was cut into three, 5 cm sections using a band saw. Each section—defined here as top, middle, or bottom—was thawed at 4 °C overnight and then homogenized by hand, removing any mineral, inorganic, or live plant material.[182]

3.3.2   *Optimized LMW DOM extraction*

To obtain a sample most consistent with compounds found free in solution and bioavailable to both plant and microbial communities,[134, 142] the soils were extracted in triplicate (three subsamples of soil) with LC/MS-grade H_2O (pH = 5.0, 1:3 w/v) in 50 mL centrifuge tubes (VWR) at 4 °C on a standard orbital shaker (VWR, Model 1000) at ~ 120 rpm for 1 h, resulting in three extracts per depth (9 total) to be analyzed by nanoLC/MS. Three controls were also prepared by adding LC/MS-grade H_2O to centrifuge tubes with no soil to undergo the same extraction procedure. Extracted soils and controls were centrifuged (Eppendorf Centrifuge 5804 R) at 4 °C and 4500 rpm for 15 min and the supernatant was then transferred to pre-rinsed centrifugal filter units (Amicon Ultra, 3 kDa) for concentration. The filtered extracts were evaporated down to 0.5 mL (12x

concentration) in a Thermo Savant SC210A SpeedVac Concentrator and separated into two 0.25 mL aliquots. One aliquot was further evaporated to near-dryness and brought back up to 0.25 mL in 95:5 (v/v) acetonitrile:water, creating one organic and one aqueous aliquot per sample for analysis by HILIC and RP, respectively.

3.3.3   Instrumentation

Here, the ESI source capillary temperature and voltage were optimized to 225 or 275 °C and 2.2 or 2.8 kV, for negative- or positive-ion mode, respectively. Full precursor (MS^1) scans were acquired in centroid mode at a resolving power of 30,000 over a mass range of 50 – 1000 m/z. Fragmentation data were collected to provide a third dimension for annotation (RT, MS^1, and MS^2) and structural information to help eliminate candidates from multiple database hits. Collision-induced dissociation (CID) with $He_{(g)}$ was performed on the top 5 ions for each full scan at 15,000 resolving power, a 2 m/z isolation width, and an optimized 30 % normalized collision energy for fragmentation. Monoisotopic precursor ions that were selected for fragmentation were placed on a dynamic exclusion list for two minutes and a charge state rejection of doubly-charged precursors was also enforced to improve detection and isolation of low abundant or coeluting small molecules. Two microscans were averaged for every full MS^1 and MS^2 spectrum to help reduce spectral complexity. Accurate m/z values were determined to four decimal places. Mass calibration was performed every two days to control for instrument drift using a mixture of caffeine, MRFA and Ultramark 1621 in ACN, MeOH, and acetic acid for positive-ion mode and a mixture of sodium dodecyl sulfate, sodium taurocholate,

and Ultramark 1621 in ACN, MeOH, and acetic acid for negative-ion mode (Pierce, Thermo Fisher Scientific).

3.3.4 *Statistical analyses*

Peak areas were $\log_2$-transformed, standardized to the dry weight of soil extracted, and normalized by LOESS and median-centering adjustments across the global dataset within the freely-available InfernoRDN software (see Chapter 2).[157] Student's t-test was used to perform pairwise comparisons between LWM DOM abundances at each depth (top, middle, or bottom) to identify the features that varied significantly ($\log_2$ FC > 1.5, p-value < 0.05) with depth. Features having a null abundance value in their triplicate were imputed with random numbers from a normal distribution. The mean and standard deviation were optimized to simulate abundance values below the noise level (width = 0.3, shift = 1.8).

3.4 Results and discussion

The goal of this work was to establish a sensitive, high-throughput, untargeted approach to detect, quantify (relative), and annotate variations in LMW DOM availability across space in Arctic soil. A preliminary analysis of Arctic soil water by RP-MS revealed that although some compounds were retained effectively, eluting later in the run, a majority (~80 %) of the most abundant ions (intensity > 5.0E4) were observed with minimal retention (RT < 2 min), and a maximum molecular weight of ~600 Da (Figure 15). This is consistent with the emerging view that much of dissolved soil organic matter is comprised of plant- or microbial-derived LMW (< 1000 Da) compounds[183] that are often polar and therefore not well-retained by RP.
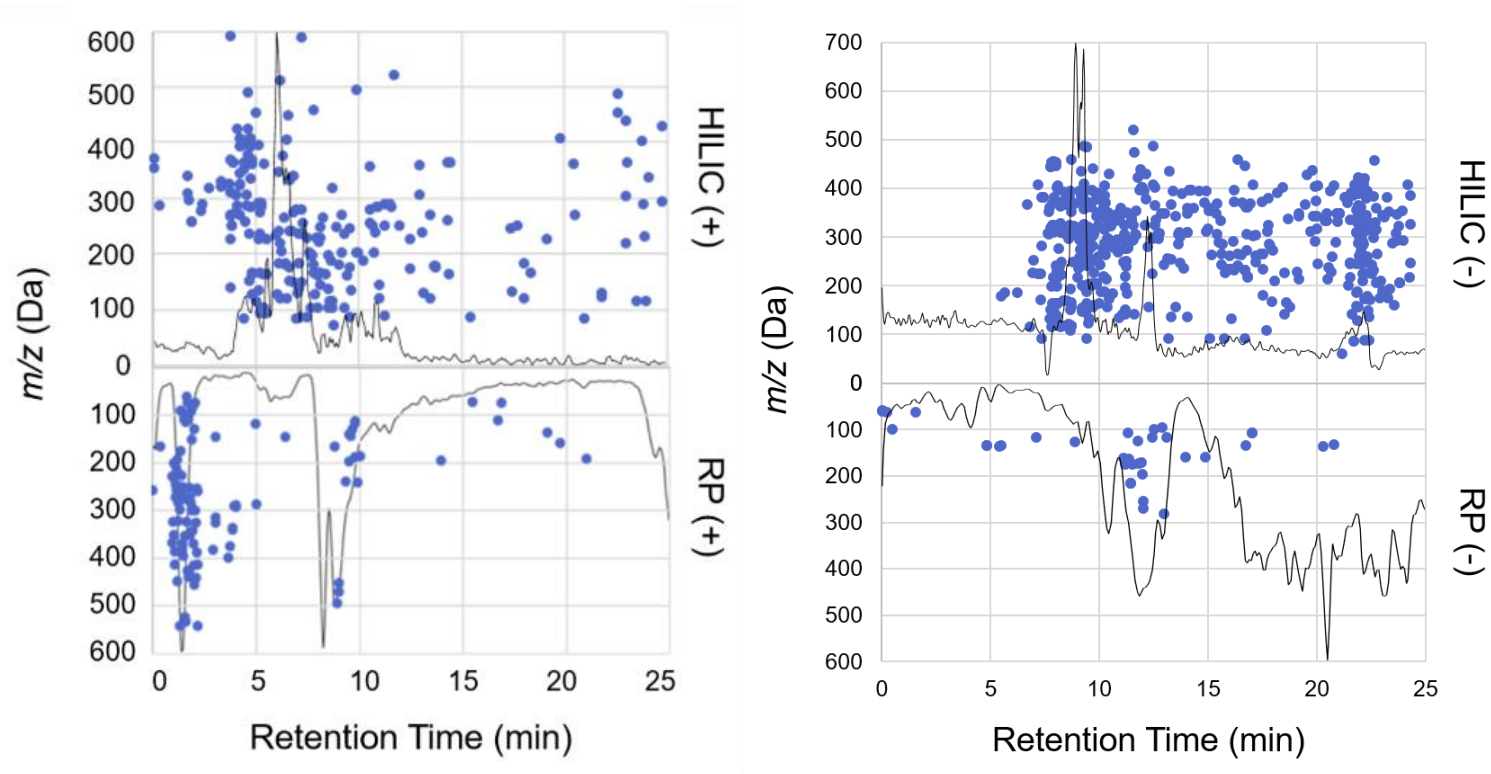
Figure 15: Scatter plots of features detected (intensity > 1.0E4, +/- 0.005 *m/z*) in a single soil water extract and the elution profiles for HILIC (top) and RP (mirrored bottom) in positive-ion mode (left) and negative-ion mode (right)

Contrasting separation profiles of LMW DOM compounds on each LC phase and polarity can be observed. Each marker matches to a *m/z* and retention time. The corresponding normalized base peak chromatograms are overlaid on top to show a typical elution profile for each LC condition and display trends between *m/z* and RT.

To enable characterization and expand coverage, we examined triplicate aqueous extractions, to mimic native soil-water chemistry, and then evaluated four nano - LC / MS analysis conditions—HILIC (+), HILIC (-), RP (+), and RP (-). Each step of the final workflow (Figure 16) was optimized to maximize throughput, enhance the signal strength of low abundant analytes, and minimize introduction of non-analyte signals which complicate annotation. The optimized approach was evaluated based on the reproducibility, separation power, and both the qualitative and quantitative performance when applied to triplicate extracts from three depths—top (samples 1-3), middle (samples 4-6), and bottom (samples 7-9)—along the organic horizon of a soil core obtained from an Alaskan Arctic landscape.

3.4.1 *Optimization of hydrophilic interaction chromatography*

Given that most LC/MS-based metabolomics analyses have used RP, were carried out at the macro-scale, or have been applied in alternate sample matrices,[138] optimizing and evaluating the nano-HILIC conditions for the separation of LMW compounds from soil water was first required. Here, we chose to exploit a zwitterionic, polymer-based HILIC material (ZIC-pHILIC) that has demonstrated improved reproducibility over other HILIC phases, and a higher tolerance for both acidic and alkaline conditions (pH range 2-10), enabling a multiple ionization strategy to be employed.[184] Optimization was carried out using a mixed standard of fifteen LMW organic compounds of varying sizes and chemical properties (Table 4).
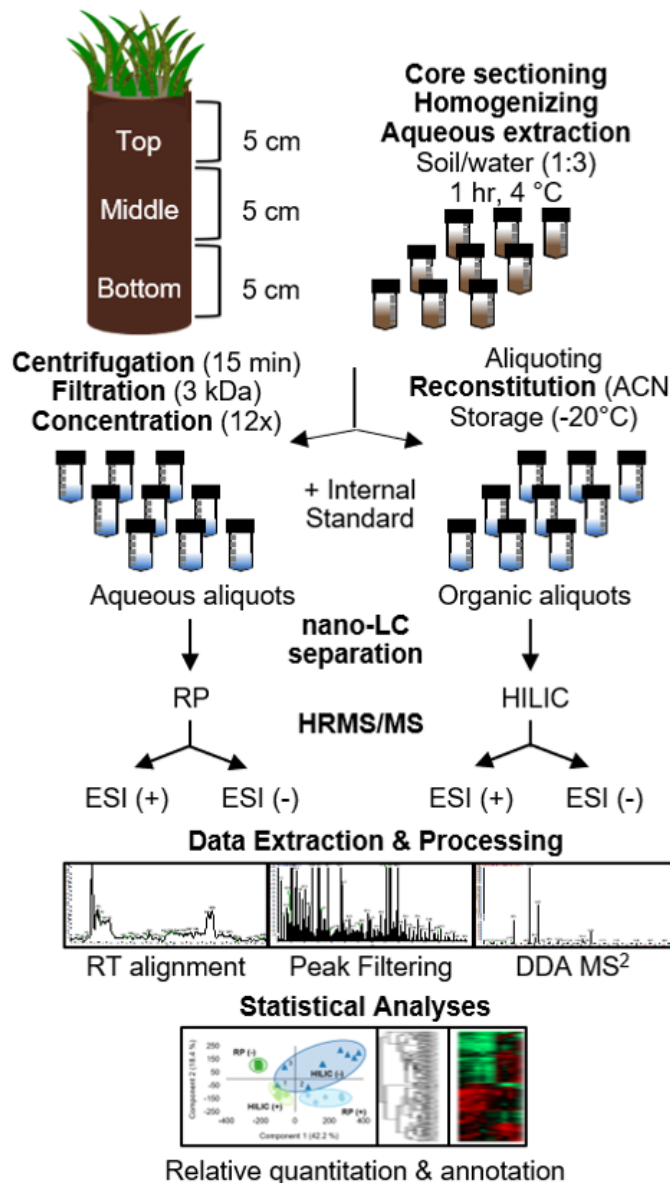
Figure 16: Schematic of the untargeted exometabolomics approach developed and applied in the present study for the analysis of LMW DOM from Arctic soil water extracts

After the filtration step, triplicate extracts for each section of the core (n = 9) were split and handled separately. The resulting concentrated aliquots (18 samples) were run on two LC phases and in two MS polarities, resulting in four analytical conditions per sample.

Table 4: List of authentic standards, low molecular weight organic compounds, used to evaluate untargeted, high-resolution mass spectrometry technique; data collected using LTQ-Orbitrap Velos Pro mass spectrometer

Compound	Class	Formula	Monoisotopic mass	Observed ion [M+H]$^+$	Mass accuracy (ppm)
Urea	Osmolyte	CH_4N_2O	60.0318	61.0393	5.65
Cytosine	Nucleobase (pyrimidine)	$C_4H_5N_3O$	111.0427	112.0497	7.55
Betaine	Osmolyte	$C_5H_{11}NO_2$	117.0784	118.0856	5.59
Adenine	Nucleobase (purine)	$C_5H_5N_5$	135.0539	136.0610	5.72
Ectoine	Osmolyte	$C_6H_{10}N_2O_2$	142.0737	143.0806	6.36
Lysine	Amino acid (basic)	$C_6H_{14}N_2O_2$	146.1049	147.1120	5.50
Glutamic Acid	Amino acid (acidic)	$C_5H_9NO_4$	147.0526	148.0596	5.67
Methionine	Amino acid (nucleophile)	$C_5H_{11}NO_2S$	149.0505	150.0575	5.54
Arginine	Amino acid (basic)	$C_6H_{14}N_4O_2$	174.1111	175.1179	6.04
N-acetyl glucosamine	Amino sugar	$C_8H_{15}NO_6$	221.0894	222.0968	1.87
Tetraglycine	Peptide	$C_8H_{14}N_4O_5$	246.0958	247.1019	7.29
6-methyl amino purine riboside	Nucleoside	$C_{11}H_{15}N_5O_4$	281.1118	282.1179	6.33
Tyrosine-Phenylalanine	Dipeptide	$C_{18}H_{20}N_2O_4$	328.1418	329.1470	7.87
YIGSR	Pentapeptide	$C_{26}H_{42}N_8O_8$	594.3126	595.3238	5.68
Chlorophyll a[a]	Pigment/ vitamin	$C_{55}H_{72}MgN_4O_5$	892.5353	871.5478	5.24

[a]Observed ion: [M-Mg+H]$^+$

3.4.2   Sensitivity and mass accuracy

To evaluate the retention of various LMW DOM compounds on the HILIC column, their electrospray ionization efficiencies, and probe detection limits and interferences, a mixed standard curve (10 ng mL^{-1} – 10 µg mL^{-1}) was spiked into and extracted from Arctic soils at ecologically-relevant concentrations[65] and analyzed by nano-HILIC-MS. All compounds were detectable and reliably quantified (S/N > 3) at 10 ng mL^{-1} or better when extracted from the soil matrix, except for N-acetyl glucosamine and urea, which were detectable at 100 ng mL^{-1} (Figure 17).

Although each of the compounds demonstrated varying ionization efficiencies, the signal response curves exhibited a linear gain in signal over at least two orders of magnitude with an average Pearson correlation coefficient (R^2) of 0.9924 demonstrating a broad dynamic range for the detection of these analytes by this technique. On average, each of the LMW DOM standards was detected within 5 ppm mass error (Table 4), demonstrating the resolution and reliability of the measurement for post-acquisition peak clustering and annotation by database searching.

3.4.3   Chromatographic reproducibility

A common challenge with untargeted LC/MS-based measurements is the ability to generate reproducible chromatograms to compare across multiple samples and obtain reliable quantitative data. To monitor the performance of the HILIC and RP columns, an internal standard (10 µg mL^{-1}) was added to triplicate extracts from each of the three soil core depths (n = 9). While it has been reported that HILIC columns often suffer from more variable peak shapes and shifting retention times,[113] the RT deviation observed here, across
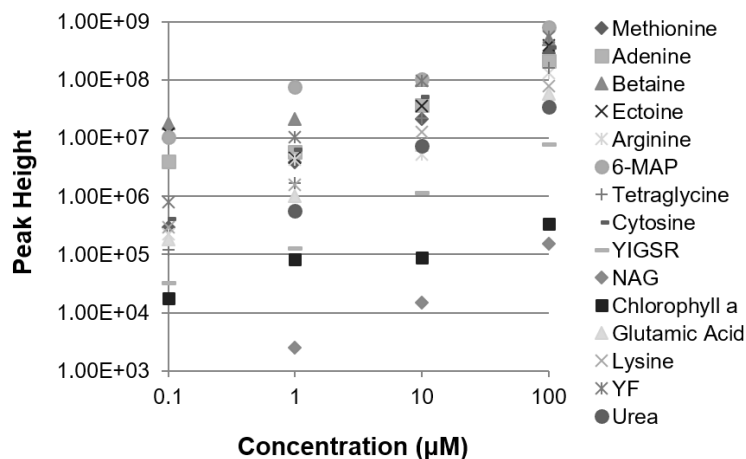
Figure 17: Signal response curves for standards spiked into and extracted from Arctic soil Standards were detected (S/N > 3) and quantified by nano-HILIC/MS in positive-ion mode, with a 20-min gradient and 1 µL injection. Average R^2 across the fifteen standards was 0.9924. Axes are shown in log scale for clarity.

all nine extracts, was < 1.8 min (CV = 12.7 %) (Figure 18), comparable to or better than the RP column. Peak areas for the internal standards also showed reasonable quantitative reproducibility among replicates (CV_{avg} < 15 %) for each LC/MS condition (Figure 19), consistent with recent studies that have also used LC/MS for untargeted metabolomic profiling in complex biological matrices.[97, 112] Notably, keeping in mind these were randomized sample analyses, there was a slightly smaller RT deviation within triplicate extractions at each depth (< 1 min, CV_{avg} = 4.8 %). These data indicate that any variations in RT were more strongly influenced by biogeochemical variation with depth than by method-derived variation, emphasizing the capacity of this technique to capture both biotic and abiotic variation (*i.e.* small pH differences, adsorption to remaining trace mineral material not removed during visual inspection/soil processing) in the availability of LMW
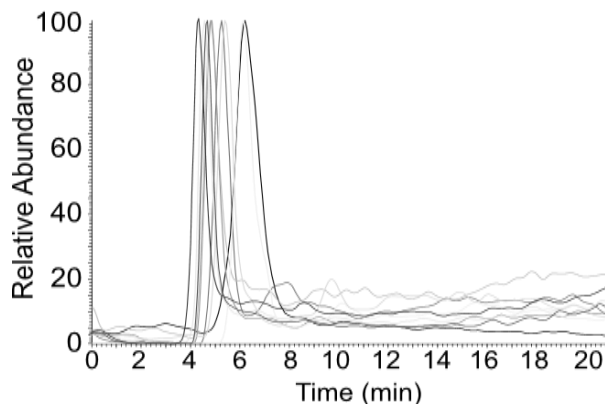
Figure 18: Normalized extracted ion chromatograms (XIC), prior to RT alignment, for the internal standard, 6-MAP, extracted from nine Arctic soil samples and detected in positive-ion mode as [M+H]$^+$ at 282.1186 *m/z* on the nano-ZIC-pHILIC column
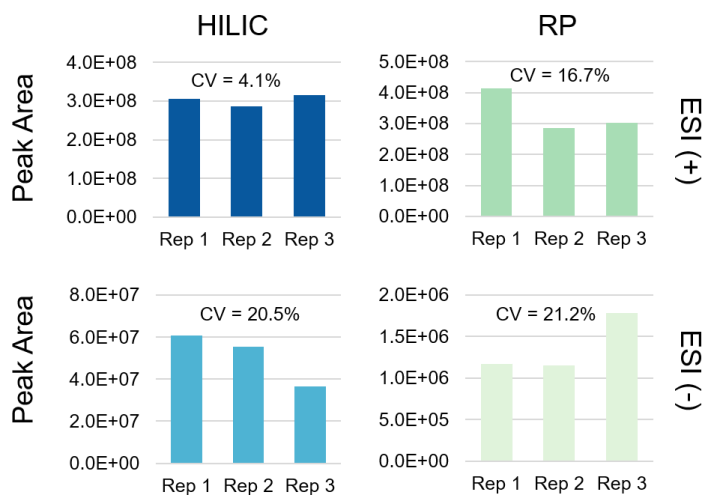
Figure 19: Integrated XIC peak areas for internal standards spiked into and extracted from triplicate soil samples (10 µM), prior to alignment or normalization procedures

6-MAP in positive-ion mode (top) and adenosine in negative-ion mode (bottom) were detected in triplicate soil water extracts on HILIC (left) and RP (right). The CV % for each triplicate is also reported (inset).

69

DOM across space. To discern to what extent adsorption to the soil phase was driving any relative quantitative differences observed with depth, we also examined the extraction efficiency of various LMW DOM standards (*i.e.* amino acids, sugars, lipids, peptides) at each of the three depths sampled. Variations in the recovery between triplicates were acceptable ($CV_{avg} < 15\%$), and also between depths ($CV_{avg} < 10\%$) with an average recovery of 88 %.

It should be noted that the HILIC column needed more time for pre-conditioning and re-equilibration to achieve a stable background, and some peak tailing was observed (Figure 18). This is likely due to competition between the primary aqueous-partitioning retention mechanism and secondary electrostatic interactions with the zwitterionic sulfobetaine group on the surface of the ZIC-pHILIC stationary phase. Nevertheless, the HILIC column demonstrated markedly improved separation and peak shape for LMW DOM analytes when compared to the RP column in this study, highlighted by the greater distribution of features eluting over the full gradient and sharper peak shapes in both positive- and negative-ion modes (Figure 15).

3.4.4 LMW DOM coverage

Expanding the number of analytes detected is central to any metabolomics study and to obtaining as unbiased and comprehensive of a measurement as possible. Across the 36 analytical runs (9 extracts, 4 LC/MS conditions), 12,924 total features were detected (Table 5). After removing artifacts, and features that resulted in zero peak area after normalization (see Materials and Methods for more detail), the total number of HQFs was 3,690. HILIC (-) detected the most with 1,705, accounting for 46 % of all HQFs observed,

followed by RP (+) with 1,462 (40 %), HILIC (+) with 438 (12 %), and finally RP (-) which detected 85 (2 %) (Table 5).

The paucity of LMW DOM analytes detected by RP (-) is likely due to poorer retention and less favorable ionization conditions. By taking each singly-charged precursor ion (+/- 0.001 m/z) to its neutral mass and analyzing the overlap between conditions (Figure 20), it was observed that HILIC (-) and RP (+) detected the most HQFs with 1,132 and 700, respectively. While these two conditions accounted for 88 % of the dataset, the four optimized techniques were highly orthogonal with just 4 % (145 features) detected by more than one condition at this high-resolution threshold (+/- 0.001 Da), illustrating the benefits of combining RP and HILIC, and positive- and negative-ion modes to expand coverage of the LWM DOM pool.

3.4.5   *Measurement depth*

In addition to expanding the number of compounds detected, an untargeted technique should be able to reliably detect both high- and low-abundant signals. This is especially true for Arctic soils, where low-abundant DOM signals could indicate a greater biological importance; in that lower concentrations may suggest a microbial preference for those substrates and that they are cycled through the soil at a faster rate, thereby contributing disproportionately to the fraction of SOM that is mineralized into CO_2 and CH_4.[59, 60] To explore the sensitivity and dynamic range of the untargeted approach developed here, we examined the proportion for which each HQF contributed to the total signal of HQFs detected by each LC/MS condition. HILIC detected more low-abundant

Table 5: LMW DOM coverage by HILIC and RP in positive- and negative-ion modes at each level of data filtering, expressed as the number of features detected across all nine soil water extracts

LC/MS Condition	All Features	High-Quality Features	Unique HQFs[a]	Abundant HQFs[b]	Varied significantly with depth[c]	Significant HQFs with MS[1] match (+/- 5 ppm)
HILIC (+)	1455	438	206	247	164	35
HILIC (-)	8343	1705	1132	257	79	14
RP (+)	1828	1462	700	202	12	8
RP (-)	1298	85	47	10	2	2

[a]Unique high-quality features observed by only one LC/MS condition, determined by examining the overlap of the neutral precursor masses (+/- 0.001 Da). [b]Abundant features were observed in at least 2 of 3 extraction replicates at each depth above an intensity threshold of 1.0E5 ion counts. [c]Abundant features with differential abundances that varied significantly ($\log_2$ FC > 1.5, p-value < 0.05) between soil core depths.
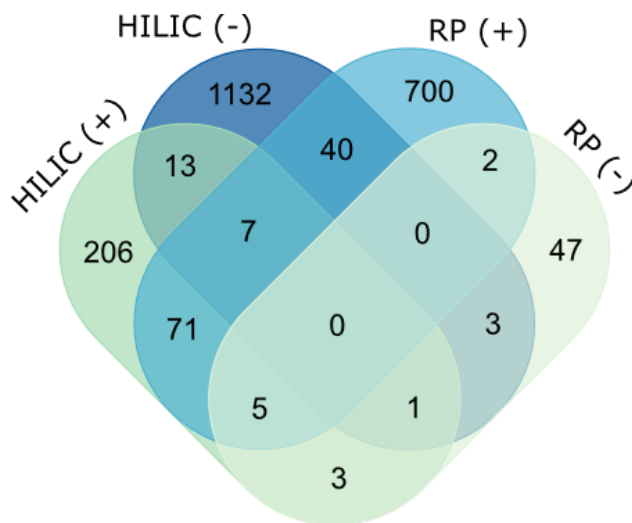
Figure 20: Overlap of HQFs detected by HILIC and RP in positive- and negative-ion MS polarities (based on MS[1] neutral mass for the corresponding [M+H]$^+$ or [M-H]$^-$ ion, +/- 0.001 Da)

features than RP, and ionization conditions leading to enhanced MS detection sensitivity.[185] For example, while only 5 features made up 50 % of the signal for RP (-), 102 different features accounted for the same proportion on the HILIC column (Figure 21).

3.4.6   Analytical reproducibility

Using a unique identifier and corresponding normalized peak area for each HQF, we evaluated the reproducibility of the untargeted measurement across extraction replicates using PCA to visualize the overall variation. When comparing the nine samples and three controls for each LC/MS condition, a strong separation was observed (Figure 22) providing additional evidence that the variation observed in the LMW DOM profiles was nonsystematic, but instead related to biogeochemical variation with depth.

PCA also revealed separation between the four LC/MS conditions (Figure 23) further demonstrating their orthogonality. HILIC (-), which detected the highest number of HQFs, showed the most variation across the nine extractions, while RP (-), which detected the fewest, showed the least amount of variation. Interestingly, the three extraction replicates within the HILIC (-) dataset that stood out from the other six, clustered closer to the other three LC/MS conditions and corresponded to samples 1-3 from the top section of the core. These data suggest that at the top of this organic horizon, there may exist a common set of abundant, amphiphilic compounds that ionize in both MS polarities, that do not get transported deeper into the organic profile.

Overall, the number of features detected by the four LC/MS conditions and the reproducibility of the untargeted measurements across extraction replicates demonstrates
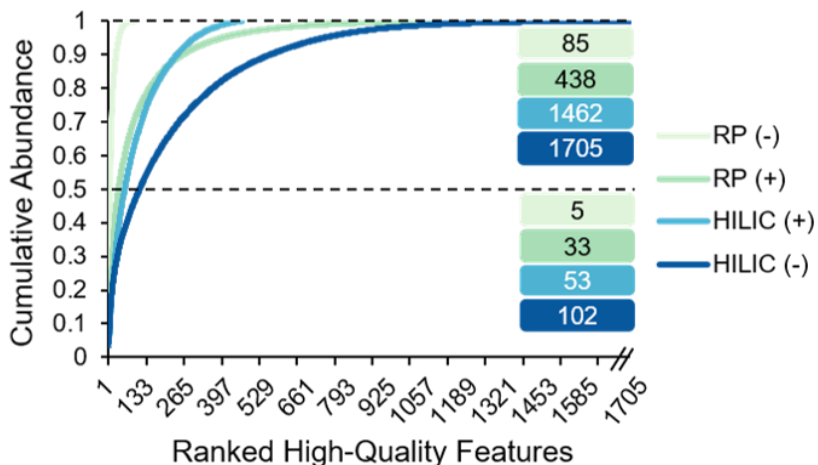
Figure 21: High-quality features ranked by abundance (1 = most abundant, 1705 = least abundant) and the relative contribution of each to the cumulative abundance

The number of LMW DOM features detected by each LC/MS condition accounting for half and the total cumulative abundance are reported demonstrating the varying depths of measurement among the LC/MS conditions evaluated.
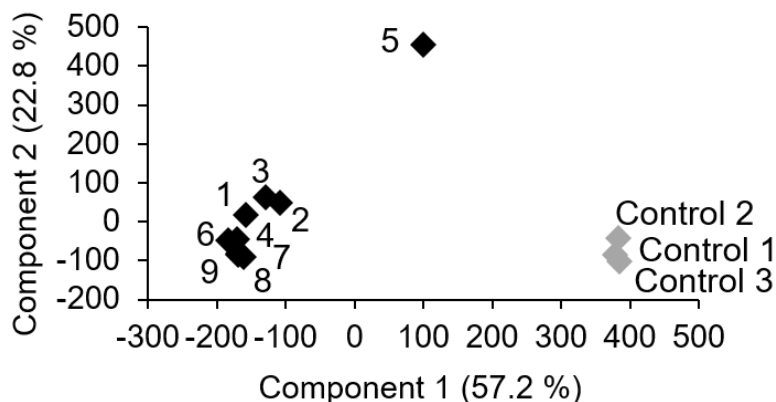
Figure 22: Example PCA of HILIC (-) dataset that used unique identifiers and peak areas to analyze the variation between features observed in the nine soil extracts and three controls, demonstrating a strong separation between LMW DOM analytes and artifacts

74
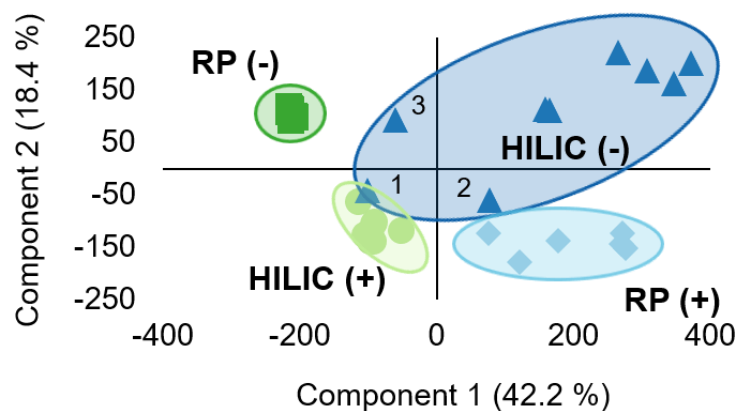
Figure 23: PCA of HQFs detected in the nine soil water extracts by each of the four LC/MS conditions evaluated

Dark blue triangles, HILIC (-); dark green squares, RP (-); light blue diamonds, RP (+); and light green circles, HILIC (+)

the robustness of the workflow developed here. Substantially more information (60 % more features) was obtained by integrating HILIC and negative-ionization mode, emphasizing the complementarity of the optimized LC/MS conditions and the ability of this untargeted technique to expand coverage of LMW DOM in these complex, organic-rich soils.

3.4.7   Application of untargeted approach to evaluate relative variations in LMW DOM availability with depth

After filtering the data to identify the *abundant HQFs* (see Chapter 2), HILIC was found to have detected a total of 247 and 257 features in positive- and negative-ion modes, respectively, while RP detected 202 in positive-ion mode and 10 in negative-ion mode (Table 5). RP (-) had less favorable mobile phase conditions and more variable chromatography which likely led to weaker ionization, lower intensities, and fewer

reproducible features. By examining a PCA for each condition separately, we found that even though the soil core represented a single horizon (organic) and would be represented as such in most biogeochemical models, the untargeted approach evaluated here revealed a fine spatial heterogeneity along the length of the horizon (Figure 24). However, instead of separating into three distinct groupings as one might expect based on our operationally-defined depths, only two groups emerged, suggesting this seemingly-homogenous organic horizon would more accurately be described as having two distinct layers, indicated by measurable differences in the LMW DOM profiles due to biogeochemical variation.

To visualize more detailed patterns of LMW DOM availability along the length of the core, hierarchical clustering using heatmaps was performed on the abundant HQFs detected by each LC/MS condition. An example of this is shown in Figure 25 using the HILIC (+) dataset. Differences in the normalized peak areas were especially apparent for two clusters that either increased or decreased from the top to the bottom of the core (Figure 25a), demonstrating the ability of the exometabolomics approach to detect variations in the LMW DOM pool between replicates and across space in soil. In addition, to generate a list of ecologically-relevant features for annotation, we identified which abundant HQFs varied significantly ($\log_2$ FC > 1.5, p-value < 0.05) with depth by t-test. The total number of features that met these criteria for each LC/MS condition are reported in Table 5. HILIC (+) and (-) detected the highest number of differentially-abundant LMW DOM features with 164 and 79, respectively, while the RP conditions detected 14 in total, demonstrating that the conservative thresholds applied here helped ensure a robust measurement of statistical significance between depths.
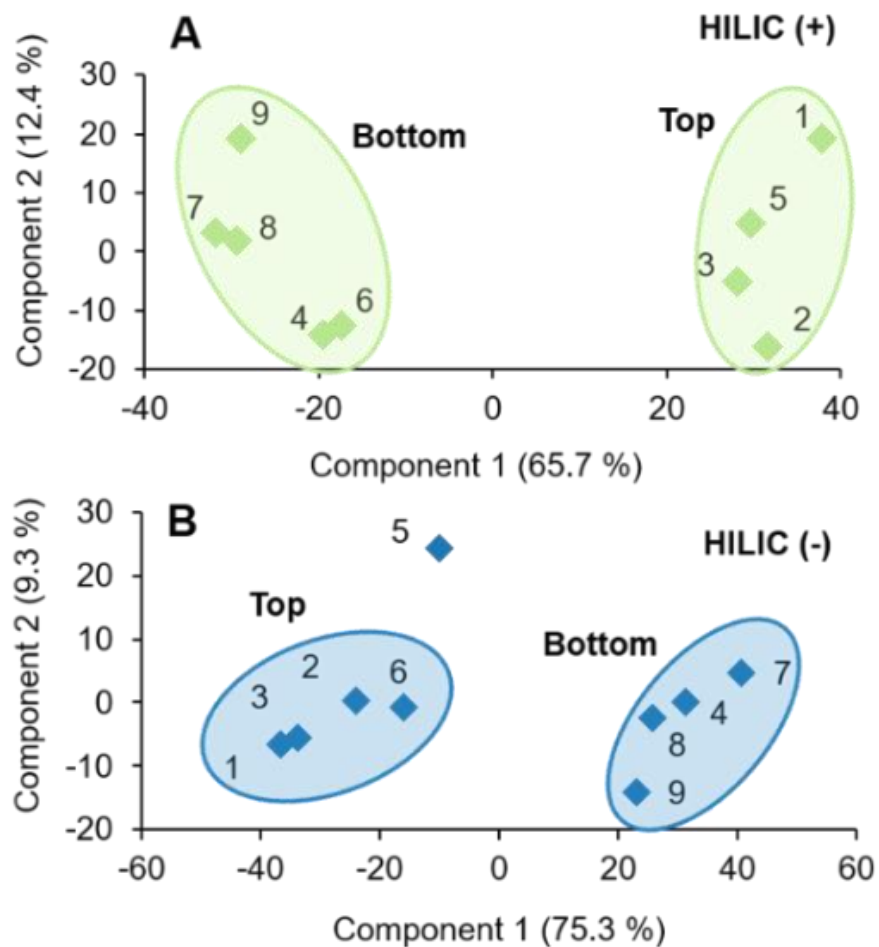
Figure 24: PCA of HQFs detected in soil water extracts analyzed by (a) HILIC (+) and (b) HILIC (-) demonstrating the sensitivity of the untargeted technique to detect subtle variations in LMW DOM with depth in these organic-rich soils
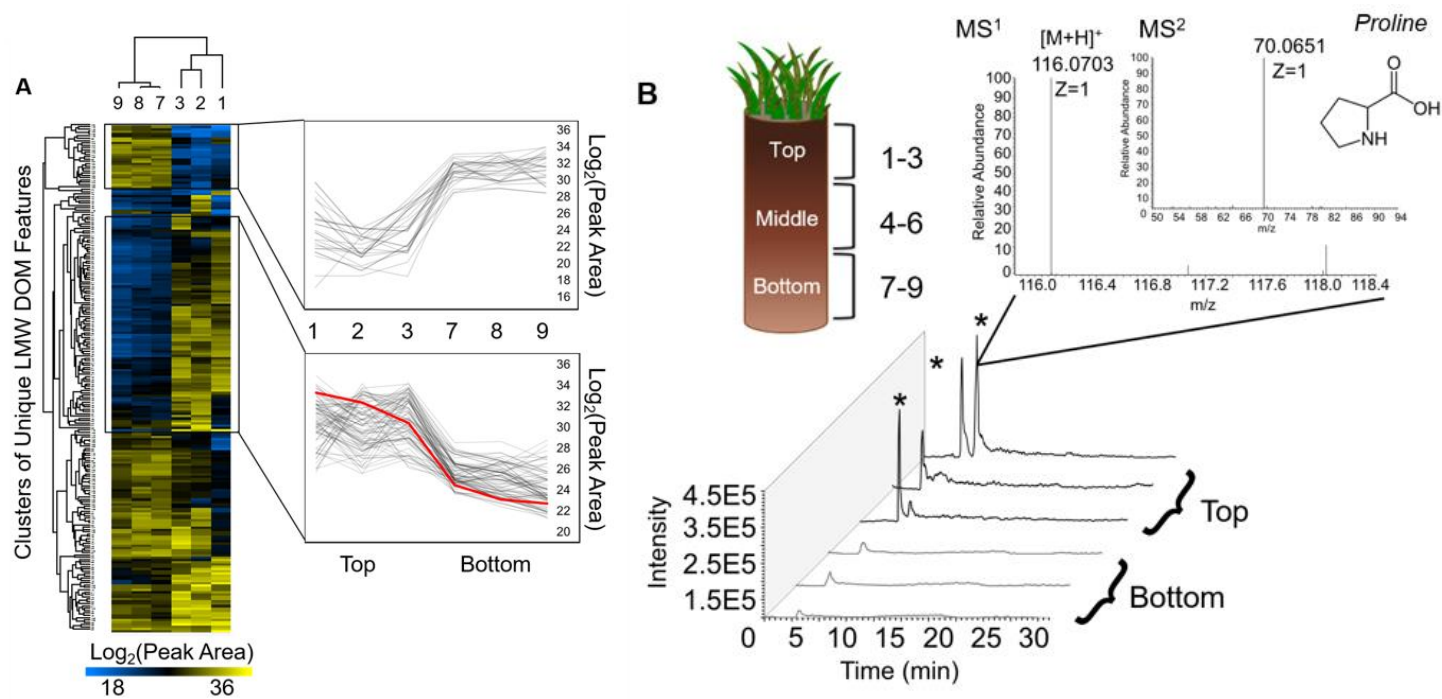
Figure 25: (a) Heatmap, or two-way hierarchically-clustered dendrogram of unique IDs and normalized log$_2$ peak areas for each differentially-abundant HQF detected by HILIC (+) with two clusters of differentially-abundant features called out (inset) (b) Cross-sectional diagram of the soil core with sample IDs and stacked XICs for feature highlighted in red in 25a, MS^1, and MS^2 spectra (insets) for a feature (116.0703 m/z) detected reproducibly by HILIC (+) at RT 6.1 min

The features that varied consistently and significantly with depth were searched against multiple freely-available online databases using high-mass accuracy (< 5 ppm) MS[1] and MS[2] measurements. When compounds matched to multiple database hits, possible matches were examined in an iterative approach by comparing the experimental fragmentation pattern with available data (Appendix B). One example of this is highlighted in Figure 25b. The feature eluted in the void volume on the RP column but was retained (RT 6.1 min) and detected (intensity $> 1.0E5$) by the HILIC column further supporting the use of dual-chromatographic separations for the analysis of LMW DOM from soil. The feature was detected in positive-ion mode ($[M+H]^+ = 116.0703$ *m/z*) reproducibly across replicates (CV = 3.01 %) and decreased significantly (4-fold $\log_2$ change, p-value < 0.05) with depth. The MS[1] accurate mass matched to multiple hits in the MMCD and HMDB databases but was putatively identified as proline by comparing the MS[2] spectrum (Appendix B) to available data in MassBank. Proline is an amino acid and osmolyte that accumulates in microorganisms and plants to help protect against stresses such as the drying and rewetting of soils.[186, 187] That it was detected appreciably in the extracellular matrix in these soils that were collected from a saturated, low topographical area (*i.e.* not drought stressed), may suggest that it had accumulated due to an increase in protease activity coupled with *reduced* uptake by plants/microbes, or enhanced exudation of excess proline from plant and microbial communities possibly due to alkaloid/salt stress.[188, 189] The decrease in this metabolite with depth may indicate that it is immediately taken up by the microbial community. Follow-up targeted analyses with labeled-proline and microbial community composition measurements for example could be carried out to monitor fluxes

and determine which of these mechanisms is dominating under similar conditions. This example demonstrates the capabilities of this untargeted, hypothesis-generating approach at identifying hotspots of biogeochemical variation for further analysis.[95] A full list of the putative identifications that were annotated in this way, within an average mass error of 3.3 ppm, can be found in Table 6 below.

Of the HQFs that consistently and significantly varied between depths, 59 (23 %) were annotated by database matching and 198 (77 %) were unmatched, highlighting a critical advantage of our approach—the ability to detect previously uncharacterized compounds that vary across space due to some biogeochemical process, thus providing targets for further inquiry. For example, one unmatched feature was retained by HILIC (-), detected reproducibly across replicates (CV < 5 %) at RT 22.7 min with an accurate mass of 281.1440 m/z and was found to increase significantly (7-fold, p-value < 0.0007) with depth. Analyzing the high-mass accuracy fragmentation data (Appendix B), neutral losses of 43.9897 m/z, 18.0106 m/z, and 14.0155 m/z were observed; likely a carboxylic acid group, water loss, and methylene group respectively, emphasizing the utility of this technique to provide structural information about unknown LMW DOM compounds. Molecular networking for untargeted -omics datasets is a growing area of research in the metabolomics community,[103, 104, 190] and leveraging high-resolution MS^2 fragmentation information like this can assist in grouping compounds based on their structural similarity.

Table 6: List of abundant HQFs that consistently ($n = 2/3$) and significantly ($\log_2$ FC > 1.5, p-value < 0.05) varied between the top and bottom of the soil organic horizon and matched to a database within +/- 5 ppm

LC/MS condition, the $[M+H]^+$ or $[M-H]^-$ ion, CV% for peak areas across triplicate extracts at both depths, Δppm from the matched compound, the predicted formula, top hit from database, which database it was detected in, and the compound class are reported. The list is sorted first by LC/MS condition, and then in order of increasing *m/z*.

LC/MS Condition	Detected *m/z*	Δppm	Predicted Formula	Top Database Hit	Database	Class (description)	CV %
HILIC (+)	72.0807	1.07	C_4H_9N	Pyrrolidine	HMDB	cyclic secondary amine; saturated heterocycle	3.30
HILIC (+)	84.0807	0.92	C_5H_9N	(+)-2,3-Dihydro-3-methyl-1H-pyrrole	HMDB	secondary amine; unsaturated aliphatic ring	4.08
HILIC (+)	86.0963	1.52	$C_5H_{11}N$	Piperidine	HMDB	heterocyclic amine	2.12
HILIC (+)	87.044	0.71	$C_4H_6O_2$	2-Butenoate;2-Butenoic acid	MMCD	carboxylic acid	3.14
HILIC (+)	90.0548	1.71	$C_3H_7NO_2$	Alanine;2-Aminopropionic acid	MMCD	amino acid	1.51
HILIC (+)	104.0705	1.04	$C_4H_9NO_2$	Beta-alanine-methyl-ester	MMCD	amino acid	2.07
HILIC (+)	115.0753	0.67	$C_6H_{10}O_2$	Gamma-hexenoic acid	LIPID MAPS	lipid	3.07
HILIC (+)	116.0705	0.93	$C_5H_9NO_2$	D-Proline;L-Proline*	HMDB	amino acid; osmolyte	3.01
HILIC (+)	120.0807	0.64	C_8H_9N	Indoline*	MMCD	aromatic heterocycle; unsaturated; bicyclic	3.54
HILIC (+)	132.1018	0.81	$C_6H_{13}NO_2$	Alloisoleucine*	MMCD	amino acid	1.42
HILIC (+)	138.0548	1.22	$C_7H_7NO_2$	Benzhydroxamic acid	MMCD	aromatic	2.41
HILIC (+)	146.0599	0.95	C_9H_7NO	Quinolin-4-ol	MMCD	monohydroxy quinoline; alcohol	2.73

Table 6 continued

LC/MS Condition	Detected m/z	Δppm	Predicted Formula	Top Database Hit	Database	Class (description)	CV %
HILIC (+)	162.1123	1.04	$C_7H_{15}NO_3$	N-methyl-4-hydroxy-leucine	MMCD	N-methyl amino acid	2.34
HILIC (+)	165.0697	1.11	$C_{13}H_8$	(E)-1,11-Tridecadiene-3,5,7,9-tetrayne	MMCD	hydroxy fatty acid, lipid	2.10
HILIC (+)	166.0861	0.92	$C_9H_{11}NO_2$	4-(3-Pyridyl)-butanoic acid	MMCD	aromatic carboxylic acid	2.44
HILIC (+)	167.9817	0.73	$C_3H_4O_6P$	Phosphoenol pyruvate; Phosphoenolpyruvic acid; PEP	MMCD	metabolite; ester, carboxylic acid	3.37
HILIC (+)	176.1028	0.96	$C_6H_{13}N_3O_3$	Citrulline*	HMDB	carboxylic acid, imine	2.57
HILIC (+)	182.0811	0.42	$C_9H_{11}NO_3$	Beta-Tyrosine*	HMDB	amino acid	4.21
HILIC (+)	184.0636	1.17	$C_5H_{13}NO_4S$	Choline sulfate	MMCD	quaternary amine, sulfate	2.58
HILIC (+)	188.0705	0.70	$C_{11}H_9NO_2$	N-(2,5-Dihydroxyphenyl) pyridinium*	HMDB	plant nutrient	6.07
HILIC (+)	189.1232	0.89	$C_8H_{16}N_2O_3$	Glycyl-Isoleucine	HMDB	dipeptide	3.76
HILIC (+)	204.0865	0.75	$C_8H_{13}NO_5$	N2-acetyl-alpha-aminoadipate	MMCD	dicarboxylic acid, amide	3.35
HILIC (+)	220.1178	0.70	$C_9H_{17}NO_5$	Pantothenate; Pantothenic acid;(R)-Pantothenate*	MMCD	secondary alcohol	2.13
HILIC (+)	226.9514	2.43	$C_6H_4Cl_2O_5$	2,4-Dichloro-3-oxoadipate	MMCD	dicarboxylic acid, ketone, dihalide	1.06
HILIC (+)	229.1545	0.80	$C_{11}H_{20}N_2O_3$	Leucyl-Proline	HMDB	dipeptide	2.51
HILIC (+)	238.092	0.58	$C_8H_{15}NO_7$	Fructoseglycine	MMCD	sugar, amino acid	6.15
HILIC (+)	251.076	0.55	$C_9H_{14}O_8$	4,6-O-(1-carboxyethylidene)-beta-D-glucose	MMCD	sugar	3.85
HILIC (+)	251.0761	0.18	$C_9H_{14}O_8$	(4AR,6R,7S,8R,8AS)-hexahydro-6,7,8-trihydroxy-2-methyl pyrano[3,2-D][1,3]dioxine-2-carboxylic acid	MMCD	lipid	3.58

Table 6 continued

LC/MS Condition	Detected m/z	Δppm	Predicted Formula	Top Database Hit	Database	Class (description)	CV %
HILIC (+)	261.144	1.75	$C_{11}H_{20}N_2O_5$	(E)-N-6-[3-carboxy-1-(hydroxy methyl) propylidene]-L-lysine	MMCD	peptide	6.98
HILIC (+)	265.1434	0.12	$C_{15}H_{20}O_4$	4-Hydroxy dehydromyo-porone	HMDB	aromatic, mono terpenoid	2.56
HILIC (+)	304.1014	4.23	$C_{12}H_{17}NO_8$	Gynocardin	HMDB	phyto-chemical	2.74
HILIC (+)	365.1564	2.60	$C_{14}H_{24}N_2O_9$	N-Acetylmuramoyl-Ala;N-Acetyl-D-muramoyl-L-alanine	MMCD	amino acid	2.77
HILIC (+)	453.2091	1.75	$C_{24}H_{30}F_2O_6$	8-isobutanoyl-neosolaniol	MMCD	sesquiterpene mycotoxin	2.45
HILIC (+)	591.3864	4.65	$C_{34}H_{54}O_8$	Lasalocid A	LIPID MAPS	lipid	3.92
HILIC (+)	635.4124	4.62	$C_{36}H_{58}O_9$	Maslinic acid	HMDB	triterpene saponin	2.30
HILIC (-)	207.0333	0.15	$C_7H_{12}O_5S$	3-(2'-methylthio) ethylmalic-acid	KEGG	plant metabolite; natural pesticide	3.62
HILIC (-)	219.1021	2.56	$C_{13}H_{16}O_3$	Ethyl 2-benzylacetoacetate*	HMDB	beta-ketoacid, plant metabolite	4.12
HILIC (-)	227.1074	1.54	$C_{10}H_{16}N_2O_2$	Pyroglutamyl-valine*	HMDB	acidic dipeptide	1.74
HILIC (-)	229.1239	2.19	$C_{15}H_{18}O_2$	8,12-Epoxy-4(15),7,11-eudesmatrien-1-one*	HMDB	metabolite; sesqui-terpenoid	3.83
HILIC (-)	241.1231	1.26	$C_{16}H_{18}O_2$	4,4'-(Butane-1,1-diyl)diphenol;1,1-Bis(4-hydroxy phenyl)butane	MMCD	metabolite	2.56
HILIC (-)	263.0968	1.62	$C_{12}H_{16}N_4OS$	2,6-diamino-8-propylsulfanylmethyl-3H-quinazoline-4-one	MMCD	microbial metabolite	4.17
HILIC (-)	265.0759	5.39	$C_{16}H_{12}NO_3$	Ungeremine	MMCD	plant metabolite; alkaloid; osmolyte; bactericide	6.49

Table 6 continued

LC/MS Condition	Detected m/z	Δppm	Predicted Formula	Top Database Hit	Database	Class (description)	CV %
HILIC (-)	271.1231	3.59	$C_{19}H_{16}N_2$	Sempervirine	MMCD	aromatic, amine	3.46
HILIC (-)	287.0947	7.63	$C_{16}H_{16}O_5$	Alkannin	MMCD	plant metabolite	1.65
HILIC (-)	293.1442	4.67	$C_{13}H_{26}O_5S$	Heptyl 1-thiohexo pyranoside*	MMCD	sugar; heteroatom	1.36
HILIC (-)	311.0816	3.23	$C_{14}H_{26}Cl_2O_2$	Methyl dichlorotri-decanoate	LIPID MAPS	lipid; fatty acid	4.54
HILIC (-)	351.1502	7.02	$C_{21}H_{22}NO_4$	Palmatine;5,6-Dihydro-2,3,9,10-tetramethoxy dibenzo [a,g]quinoliziniu m	MMCD	metabolite	4.44
HILIC (-)	371.1039	5.25	$C_{19}H_{20}N_2O_4S$	2-(1,3-dioxo-1,3-dihydro-2H-isoindol-2-YL) ethyl-4-(4'-ethoxy [1,1'-biphenyl-4-YL)-4-oxbutanoic acid	MMCD	metabolite	3.77
HILIC (-)	457.1309	9.33	$C_{20}H_{26}O_{12}$	3-O-a-L-Arabinofuranosyl -D-xylose, 2-O-(4-Hydroxy-3-methoxy cinnamoyl)*	HMDB	plant sugar	3.57
RP (+)	60.0444	0.19	C_2H_5NO	Aminoacet-aldehyde*	MMCD	alkylamine	0.76
RP (+)	85.0284	0.09	$C_4H_4O_2$	4-Hydroxy-2-butenoic acid gamma-lactone	HMDB	organic acid	2.00
RP (+)	101.0709	0.35	$C_4H_8N_2O$	N-nitroso-pyrrolidine	HMDB	basic heterocycle; secondary amine	1.27
RP (+)	148.0602	1.66	$C_5H_9NO_4$	Glutamic acid	KEGG	metabolite	2.41
RP (+)	176.103	0.17	$C_6H_{13}N_3O_3$	D-Citrulline*	MMCD	amino acid	4.18
RP (+)	212.1644	0.50	$C_{12}H_{21}NO_2$	Elaeokanine C	PubChem	alkaloid	2.95
RP (+)	216.1958	0.07	$C_{12}H_{25}NO_2$	12-amino-dodecanoic acid	LIPID MAPS	carboxylic acid, amine	3.30

Table 6 continued

LC/MS Condition	Detected *m/z*	Δppm	Predicted Formula	Top Database Hit	Database	Class (description)	CV %
RP (+)	226.1285	0.06	$C_8H_{19}NO_6$	5-deoxy-5-[(1S)-1-hydroxyethyl]amino-D-glucitol	MMCD	sugar	3.33
RP (-)	195.0512	0.86	$C_6H_{12}O_7$	L-Gulonate;L-Gulonic acid;Gulonate*	MMCD	sugar-derivative	11.2
RP (-)	269.2493	2.48	$C_{17}H_{34}O_2$	15-methyl palmitic acid	LIPID MAPS	lipid	4.93

*Indicates experimental and database MS^2 information reported in the Appendix B.

Classes of compounds annotated ranged in polarity and aromaticity, from plant and microbial metabolites to organic acids, osmolytes, sugars, lipids, and simple peptides (Table 6), demonstrating the chemical diversity of LMW DOM in Arctic soil water detected by the optimized platform. As with any untargeted approach, the number of features annotated depends on the level of curation of each database, and the features listed here therefore do not represent all LMW DOM molecules that can be annotated by the described technique. It's important to note that our aim was not to identify each feature detected but instead to evaluate the approach in this new and complex matrix, demonstrate the value of the untargeted approach in revealing an information-rich molecular profile of LMW DOM availability in soil, and to analyze how this approach may be used to evaluate variations in those profiles across space (here, with depth). Further examination of feature clusters that varied similarly and significantly with depth would likely reveal additional biogeochemical processes impacting the availability of these compounds, but additional soil core replicates would be necessary. Follow-up targeted analyses (*e.g.* isotopic or flux analyses) could be carried out for absolute quantitation of LMW DOM analytes-of-interest or to monitor a specific metabolic pathway (e.g. methanogenesis) between conditions.

3.5   Conclusions

These results demonstrate an optimized approach for discovery-based exometabolomics in soil water extracts and for distinguishing key LMW DOM analytes for further evaluation. The optimized approach developed here was sensitive and robust, with a high tolerance for salts, and could feasibly be applied in a broad range of soils. The LC/MS conditions were highly complementary and revealed a broad diversity of small molecules in Arctic soil water extracts. Furthermore, LMW DOM profiles were reproducible and distinguishable between samples. Even subtle, but consistent and significant differences in the relative abundance of features with depth were detected using robust data mining strategies, highlighting the potential of the LMW DOM pool to provide a chemical snapshot of biological activity in soil. Thus, in this chapter, we showed that this platform is useful not only for characterizing LMW DOM, but also for quantifying relative variations in the availability of LMW DOM with depth, revealing hotspots of biogeochemical activity for further evaluation.

CHAPTER 4: UNTARGETED EXOMETABOLOMICS REVEALS BIOGEOCHEMICAL HOTSPOTS WITH VEGETATION AND POLYGON TYPE IN ARCTIC TUNDRA SOILS

The chapter presented below has been adapted from the following manuscript being prepared for submission in:

Ladd, M.P., Reeves, D., Poudel, S., Iversen, C.M., Wullschleger, S.D. Hettich, R.L. Untargeted exometabolomics reveals biogeochemical hotspots with vegetation and polygon type in Arctic tundra soils. *Environmental Science & Technology* (*in prep*).

ML's contributions included: literature review, experimental design, sample collection, sample preparation, data collection, data analysis, manuscript writing and editing. DR and SP assisted with data collection and analysis, respectively.

4.1   Abstract

Rising temperatures in the Arctic have led to rapid thawing of permafrost soils, which has had interacting effects on landscape geomorphology, hydrology, and plant and microbial communities, all of which influence the cycling of C, N, and P in these systems. Characterizing how the availability of LMW DOM correlates with these landscape-scale properties is critical to understanding how SOM chemistry may be used in predictive models of C cycling in the Arctic. Despite this, little is known about how LMW DOM varies across the Arctic landscape. In this study, we applied the optimized dual-LC, dual-polarity, nano-ESI-MS/MS approach from Chapter 3 to soil organic horizons with two contrasting aboveground landscape topographies and vegetation profiles, to yield new insights into the diversity of organic species available to Arctic plant and microbial communities and elucidate the molecular distribution of LMW DOM across these difference landscape conditions. Given that this is the first application of this technique across multiple sites, the analytical performance of the approach was first evaluated. Then, due to the large amount of data generated, a series of data mining techniques and multivariate statistical analyses were applied to reduce the dimensionality of the data, discriminate ecologically-relevant features, and evaluate compositional variations due to

polygon type or vegetation. Features that were significantly differentially-abundant between sites were further investigated and annotated using high-mass accuracy MS data for formula assignment and database searching. Characterizing LMW DOM across multiple landscape features in Arctic soils will enhance our understanding of the controls on SOM decomposition, and provide data that could help reduce uncertainty in mechanistic models of C cycling in these systems.

4.2   Introduction

Polar tundra, a primary landscape type in Arctic systems, is often dominated by characteristic features called ice-wedge polygons that form when freeze-thaw cycles physically move the soil. This creates a unique microtopography across the landscape (Figure 26) which has been shown to strongly influence hydrology, vegetation, and microbial community structure.[48, 50, 191] There are different types of polygons including low-centered polygons (LCP) that have a topographically low and generally wet center, that over time can turn into high-centered polygons (HCP) which have topographically higher and dryer centers (Figure 27).[191] These features are typically ~ 5-20 m in diameter and act as distinct, repeatable units across the landscape that are valuable for scaling up measurements, and initializing landscape model integrations.[192] Recently, numerous studies have reported a strong relationship between polygon type, vegetation, and biogeochemistry (*i.e.* inorganic ions, pH, redox potential, bulk C/N),[46, 47, 193, 194] especially in the organic-rich active layer.[195] However, LMW DOM chemistry and how it varies with polygon type or vegetation remains poorly understood. Here, we characterize LMW DOM

Figure 26: Aerial photograph of Arctic polygonal tundra landscape on the northern coastal plain of Alaska near Utqiaġvik

Source: Image courtesy of U.S. Department of Energy, Oak Ridge National Laboratory
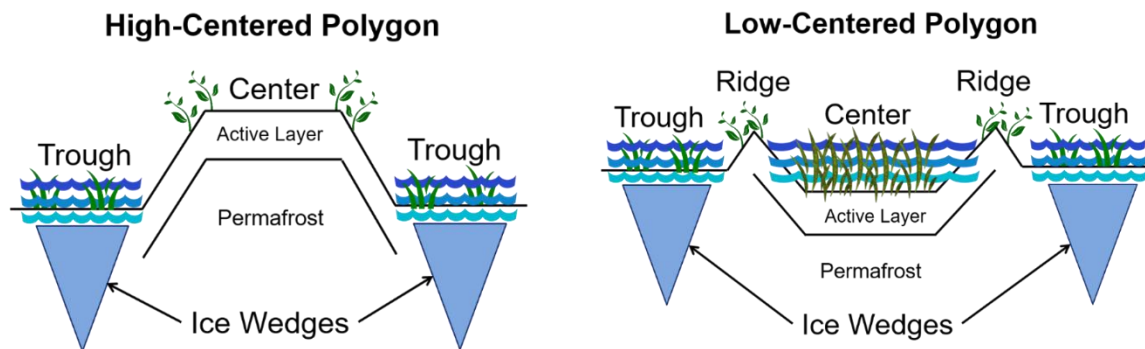
Figure 27: Cross-sectional illustrations of a (left) high- and (right) low-centered ice wedge polygon demonstrating their different microtopographies and associated variations in hydrology, vegetation, and thaw depth

in soil cores collected from the centers of an LCP and HCP, with two contrasting aboveground vegetation profiles, using the nanoLC/MS approach optimized in Chapter 3.

4.3   Experimental approach

4.3.1   *Study site and sample description*

Soil cores (n = 4, organic horizon only, 10 cm dia., ~30 cm depth) were collected from the BEO, a polygonal tundra landscape on the northern coastal plain of Alaska. To examine the relationship between polygon type, vegetation, and LWM DOM availability, two cores were collected from the center of an LCP and two from the center of an HCP, where the aboveground vegetation in one core at each site was primarily either *Carex aquatilis* or *Eriophorum angustifolium*, two dominant plant species in these systems. Due to logistical constraints, replicate cores with the same vegetation and same polygon type were not available, but triplicate samples from each core were analyzed by the LC/MS approach to ensure statistical relevance and enable a comparative analysis. The cores were collected in late-August 2014 when the active layer had reached its maximal depth (~ 34 cm).[196] There were no visible signs of cryoturbation in each horizon. The mean air temperature for this region during August is 4 °C and the mean annual precipitation is 10.74 cm.[197] Additional information about the study site and soil type has been described in detail previously.[198] The cores were shipped frozen to ORNL where they were stored at -80 °C until processing.

Each core was sectioned into three 5-cm sections, thawed, and each section homogenized by hand as described in Chapter 3, to enable the evaluation of any within-horizon variations. Live roots were removed, dried, and weighed (Table 7) to evaluate any

Table 7: Polygons soil core sample summary – TOC, TN, TC, C:N, and dry root weight results

Extract Number	Site	Polygon Type	Vegetation	Water Content (%)	g H2O/g dry soil	TOC (%)	TN (%)	TC (%)	C:N	Dry Root Weight (g)
1	A	LCP	Carex							
2	A	LCP	Carex	82.8	4.84	41.415	2.570	48.290	16.117	0.1444
3	A	LCP	Carex							
4	A	LCP	Carex							
5	A	LCP	Carex	80.3	4.07	42.624	2.228	46.966	19.131	0.1814
6	A	LCP	Carex							
7	A	LCP	Carex							
8	A	LCP	Carex	79.6	3.91	42.104	2.241	46.016	18.792	0.0778
9	A	LCP	Carex							
10	B	HCP	Carex							
11	B	HCP	Carex	73.4	2.75	41.521	2.479	46.621	16.750	0.3746
12	B	HCP	Carex							
13	B	HCP	Carex							
14	B	HCP	Carex	72.1	2.58	43.464	2.567	45.209	16.930	0.0605
15	B	HCP	Carex							
16	B	HCP	Carex							
17	B	HCP	Carex	73.4	2.76	37.334	2.1915	41.741	17.036	0.1694
18	B	HCP	Carex							
19	A	LCP	Eriophorum							
20	A	LCP	Eriophorum	85.1	5.69	34.253	1.931	47.441	17.742	1.0316
21	A	LCP	Eriophorum							

Table 7 continued

Extract Number	Site	Polygon Type	Vegetation	Water Content (%)	g H2O/g dry soil	TOC (%)	TN (%)	TC (%)	C:N	Dry Root Weight (g)
22	A	LCP	Eriophorum							
23	A	LCP	Eriophorum	83.8	5.16	35.809	2.242	47.098	15.971	0.5866
24	A	LCP	Eriophorum							
25	A	LCP	Eriophorum							
26	A	LCP	Eriophorum	76.5	3.26	38.673	2.308	43.615	16.755	0.1730
27	A	LCP	Eriophorum							
28	B	HCP	Eriophorum							
29	B	HCP	Eriophorum	75.6	3.09	39.803	2.189	47.619	18.185	1.1620
30	B	HCP	Eriophorum							
31	B	HCP	Eriophorum							
32	B	HCP	Eriophorum	73.8	2.82	39.554	2.357	46.203	16.782	0.3173
33	B	HCP	Eriophorum							
34	B	HCP	Eriophorum							
35	B	HCP	Eriophorum	70.4	2.38	41.026	2.431	44.054	16.873	0.0446
36	B	HCP	Eriophorum							

correlation between LMW DOM abundance and dry root weight. A subsample from each core section was taken to determine water content, total C and N, and total organic carbon (Table 7) using conventional techniques, described in Chapter 2.

4.3.2 Soil extraction and sample preparation

Biological replicates were obtained by extracting each core section in triplicate (n = 36, nine per core), along with three controls (extraction with no soil), using the procedure optimized for these soils described above in Chapters 2 and 3. Briefly, a single aqueous extraction (LC/MS-grade H_2O, pH = 5, 1:3 w/v, 1 hr) was employed to maintain high-throughput and obtain a sample most consistent with compounds that would be found free in soil solution and bioavailable to both plant and microbial communities.[134, 142] Same as before, extracts were centrifugal filtered (Amicon Ultra, 3 kDa, 4°C, 15 min), concentrated down (12x), and then separated into two aliquots. One aliquot was further evaporated to near-dryness and brought back up in 95:5 (v/v) ACN:H_2O, creating one organic and one aqueous aliquot per sample for analysis by HILIC and RP-LC, respectively. Extracts were stored at -80 °C until LC/MS analysis.

4.3.3 Instrumentation and LC/MS data collection

Samples and controls were thawed and prepared immediately prior to injection by adding either FA or NH_4OH (0.1 %) to help with ionization in positive- or negative-ion mode, respectively. Each sample was manually injected directly onto the columns using a 300 nL fused-silica loop, and nano-flow rates were achieved using a split-flow setup prior to the injection loop. The QCs were run every 6 injections and samples were randomized to reduce instrument-derived variation. Technical blanks representing the column re-

equilibration conditions were also run regularly to monitor background ions and carry-over between samples.

Separations were performed using the same HILIC and RP-LC phases, setup, and optimized mobile phase conditions described in Chapters 2 and 3. In addition, the same HPLC pump, mass spectrometer, and MS parameters were used for this study again resulting in four separate LC/MS analyses per sample (HILIC +/- and RP +/-, n = 144). To control for instrument drift, the mass spectrometer was externally calibrated every two days or before switching columns or polarities.

4.3.4   *Untargeted LC/MS data processing*

Raw LC/MS files were processed using the freely-available MZmine (v2.30) software.[154] Detailed descriptions of each of the modules used for peak detection, chromatogram alignment, peak list generation, and annotation can be found in Chapter 2 and screen captures of the bioinformatic workflow for an example dataset have been provided in Appendix A. The parameters used for each module in this study are listed below in Table 8. Briefly, MS^1 precursor ions that were selected for fragmentation and had an intensity above a specified noise level (S/N > 3) were added to a peak list for further analysis. Chromatograms were then built using an algorithm that searches for the same feature (MS^1 and MS^2) in both directions of the retention time within a given *m/z* and RT tolerance (+/- 2 min), resulting in a single assigned peak area. All chromatograms within each LC/MS condition were aligned across the sample set (including blanks and controls) using the RANSAC algorithm—a RT correction tool that uses a nonlinear regression model

95

Table 8: MZmine parameters used for each module applied in the analysis of the polygonal tundra soil organic horizons

Peak Detection Methods				
Mass Detection	**HILIC (+)**	**HILIC (-)**	**RP (+)**	**RP (-)**
RT window:	Auto range	Auto range	Auto range	Auto range
MS level:	1 and 2	1 and 2	1 and 2	1 and 2
Polarity:	+	-	+	-
Spectrum type:	centroided	centroided	centroided	centroided
MS1 noise level:	1.00E+04	2.00E+05	5.00E+03	1.00E+03
MS2 noise level:	5.00E+02	4.00E+02	5.00E+02	1.00E+02
MS/MS Peak List Builder				
RT window:	Auto range	Auto range	Auto range	Auto range
MS level:	2	2	2	2
Polarity:	+	-	+	-
Spectrum type:	centroided	centroided	centroided	centroided
m/z window	0.01	0.01	0.01	0.01
Time window	61 min	56 min	41 min	41 min
Peak Extender				
m/z tolerance:	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm
Min height	1.00E+04	1.00E+02	1.00E+03	1.00E+03
Peak List Methods				
Isotopic Peaks Grouper	**HILIC (+)**	**HILIC (-)**	**RP (+)**	**RP (-)**
m/z tolerance:	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm
RT tolerance:	1.0 min	1.0 min	1.0 min	1.0 min
Monotonic shape:	Y	Y	Y	Y
Maximum charge:	1	1	1	1
Representative isotope:	Most intense	Most intense	Most intense	Most intense

Table 8 continued

Duplicate Peaks Filter	HILIC (+)	HILIC (-)	RP (+)	RP (-)
m/z tolerance:	0.001 m/z or 5 ppm	0.001 m/z or 5 ppm	0.001 m/z or 5 ppm	0.001 m/z or 5 ppm
RT tolerance:	0.25 min	0.25 min	0.25 min	0.25 min
RANSAC Aligner				
m/z tolerance:	0.005 mz or 10 ppm	0.005 mz or 10 ppm	0.005 mz or 10 ppm	0.005 mz or 10 ppm
RT tolerance:	61 min	56 min	41 min	41 min
RT tolerance after correction:	20 min	30 min	20 min	20 min
RANSAC Iterations:	0 (model optimized)	0 (model optimized)	0 (model optimized)	0 (model optimized)
Minimum number of points:	25%	25%	30%	30%
Gap Filling				
m/z tolerance:	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm
RT tolerane:	61 min	56 min	41 min	41 min
Annotation				
m/z tolerance:	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm	0.005 m/z or 10 ppm
m/z vs RT balance:	0.2 min	0.2 min	0.2 min	0.2 min
Max fragment peak height:	80%	80%	80%	80%
Min MS2 peak height:	5.00E+02	1.00E+02	5.00E+02	1.00E+02
Max complex peak height:	50%	50%	50%	50%
Max relative adduct peak height:	50%	50%	50%	50%
Online databases searched:	KEGG, PubChem, HMDB, LipidMaps, PlantCyc	KEGG, PubChem, HMDB, LipidMaps, PlantCyc	KEGG, PubChem, HMDB, LipidMaps, PlantCyc	KEGG, PubChem, HMDB, LipidMaps, PlantCyc

to align chromatographic peaks across samples. Although a soft-ionization technique, electrospray ionization can create in-source fragments, adducts, or ion complexes that can complicate spectral analysis and annotation. Using the identification module in MZmine (Appendix A), each spectrum was searched for adducts, complexes, and fragments using specified RT and *m/z* thresholds (Table 8). The proportion of each LC/MS dataset identified as either adducts, complexes, or fragments did not exceed ~ 10 % (Figure 28) and can be removed from the dataset at any point in data filtering process. For the sake of evaluating the technique, here, they were not removed in order to evaluate the proportion of which may be annotated as LMW DOM metabolites by database searching as well.

4.4   Results and Discussion

4.4.1   *Evaluation of analytical performance across multiple sites*

Given that a detailed analysis of the analytical performance of this untargeted LC/MS approach in Arctic soils was conducted in Chapter 3, only a few primary figures of merit—measurement depth, reproducibility, and LMW DOM coverage—were examined here. This assisted with evaluating any methodological impacts from expanding the analysis from a single core at one location to multiple cores from different sites across the landscape. All data processing, filtering steps, and statistical analyses were conducted separately for each LC/MS condition (HILIC +/-, RP +/-) to eliminate any confounding effects such as different ionization efficiencies or noise levels for example.

Across the four conditions, 13,673 molecular species (RT, MS^1, and MS^2) were detected, aligned, and exported for data filtering and analysis (Table 9). A preliminary PCA
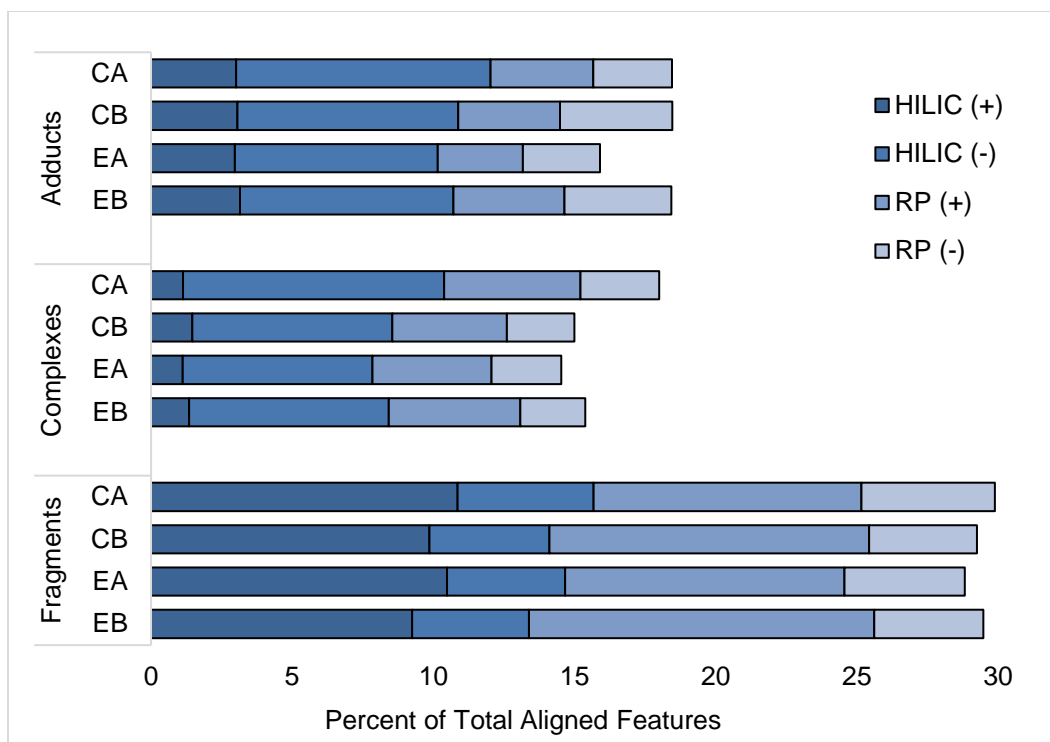
Figure 28: Percent of aligned peaks that were annotated as a possible adduct, complex, or fragment of another feature within 0.1 min and 5 ppm mass accuracy for each core grouped by LC/MS condition

C = *Carex*, E = *Eriophorum*, A = Site A or low-centered polygon, B = Site B or high-centered polygon

Table 9: LMW DOM coverage by HILIC and RP in positive- and negative-ion mode at each level of data filtering, expressed as the number of features detected across all 36 soil water extracts from 4 cores obtained from 2 polygon types and 2 species of vegetation

	HILIC (+)	HILIC (-)	RP (+)	RP (-)
Aligned peaks[a]	4686	2853	4213	1921
Features[b]	4352	2249	3655	1762
High-Quality Features[c]	3929	2170	3618	1541
Unique HQFs[d]	3414	1942	3494	1287
Abundant HQFs[e]	1966	776	1259	99
Differentially-abundant[f]	322	76	122	1
Annotated[g]	283	74	117	1

[a]Aligned peaks with same RT, MS^1, and MS^2 data from MZmine [b]After zeros and artifacts (observed in blank or control) were removed [c]Single RT, MS^1 (duplicates removed) and corresponding MS^2 spectrum [d]Number of features remaining after overlap analysis where isomers and isobars were removed [e]Observed in at least three samples across each core [f]Passed paired t-test p-value of < 0.001 and FC > 4 [g]Database match within 5 ppm, MS/MS confirmation, and biologically-relevant compound or elemental formula assignment using high-mass accuracy MS^1 data and element heuristics

for each LC/MS dataset, prior to any filtering, normalization, or statistical procedures, revealed a clear separation between blanks, controls, and samples (Figure 29), indicating that variations observed between samples were not experimentally-derived but instead due to biogeochemical variation. However, ~ 18 % of the aligned peaks were observed in nearly all the runs including the blanks and controls (Figure 30), suggesting these were background signals from the sample preparation procedures or LC/MS analyses. After removing these, as well as any zeros or duplicate features, 11,258 HQFs remained for downstream analyses (Table 9). When we plotted the frequency at which these remaining features were observed across the dataset, we noted a recurrent trend in the data where the number of features that were observed increased sharply approximately every nine samples (Figure 30), corresponding with the sample set size for each core (9 extracts). These results indicate that the data filtering protocol employed here effectively reduces the number of false positives and increases the proportion of LMW DOM analytes represented. These results also suggest that a common set of LMW DOM features exists within each core, and across all four cores, despite variations in aboveground vegetation or topography. Indeed, when we examined the overlap between the four cores for each LC/MS condition using the neutral mass for each $[M+H]^+$ or $[M-H]^-$ singly-charged precursor ion within 0.005 Da, on average there was a 37 % overlap in the features detected (Figure 31). Contrastingly, on average, 15.5 % of the features detected were found to be unique to each core indicating there was unique biogeochemical activity within each core as well, and that the optimized LC/MS approach employed here is sensitive enough to detect these subtle variations across multiple sites.
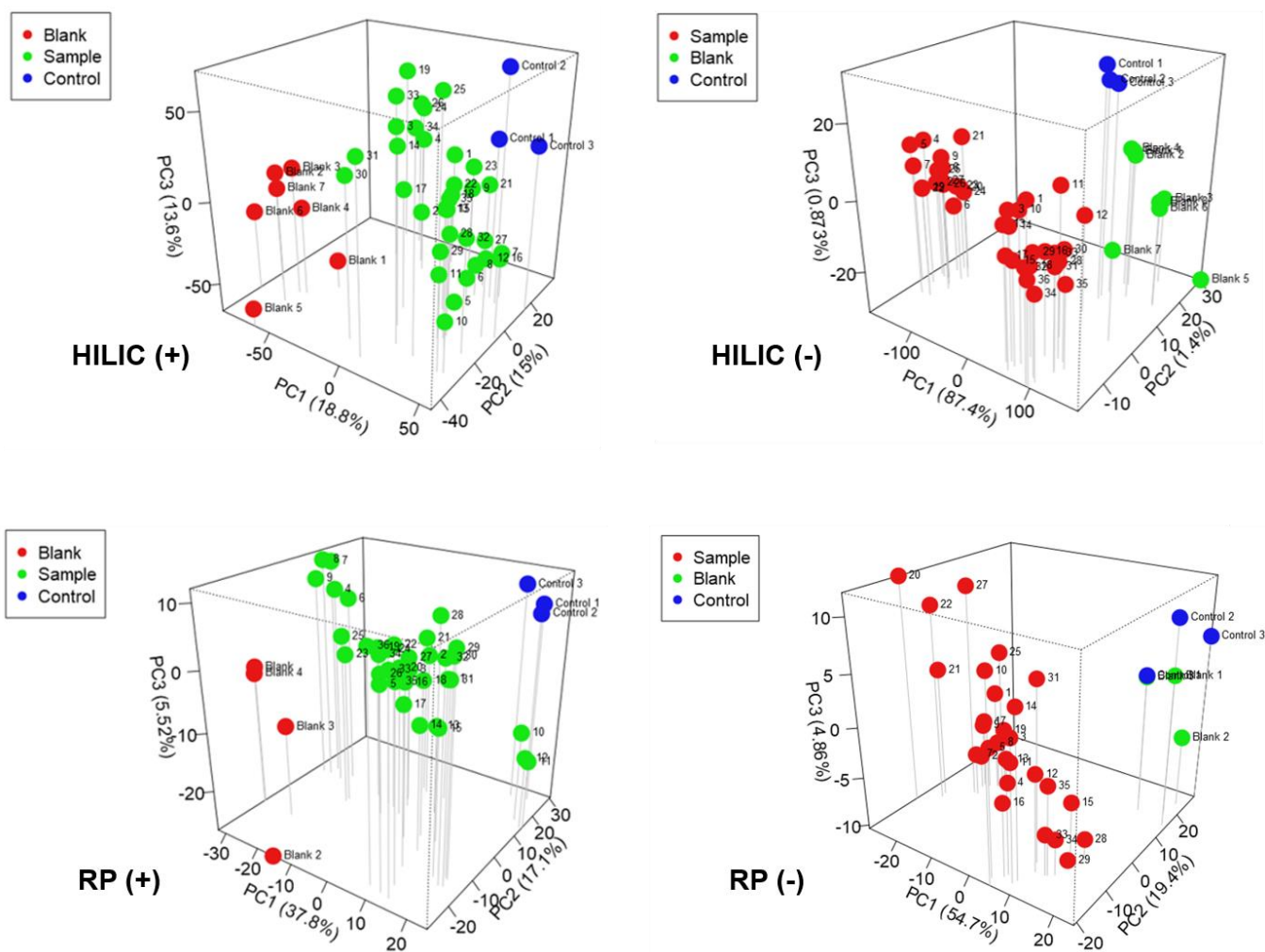
Figure 29: PCA of raw log$_2$ peak areas for blanks, controls, and samples separated by LC/MS condition
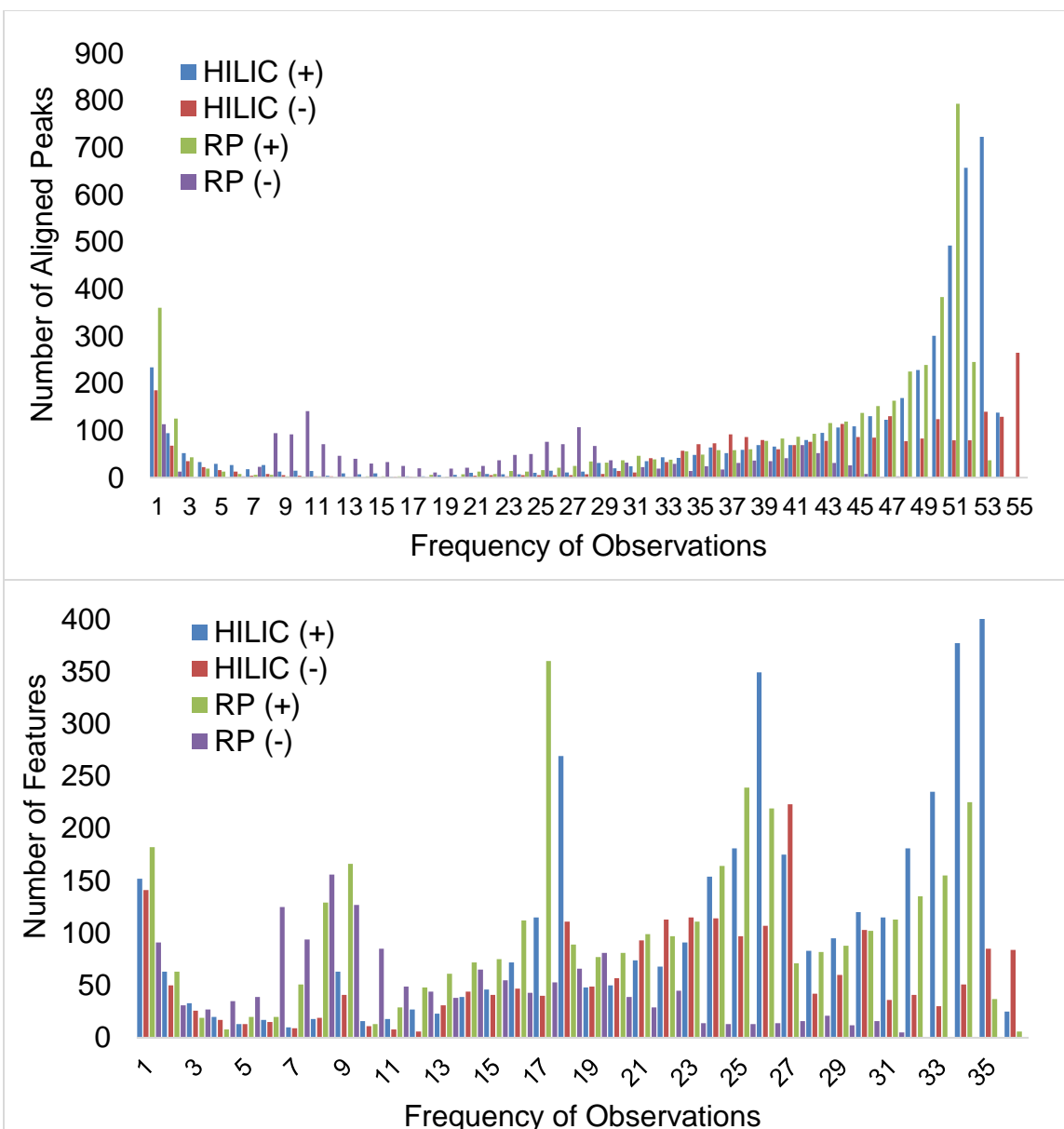
Figure 30: Histogram of the frequency of (top) observations for each aligned peak (RT, MS^1, MS^2) across the entire dataset (all 4 cores), including blanks and controls (55 total runs), prior to data filtering and (bottom) HQFs that were observed across the 36 samples after removing background peaks
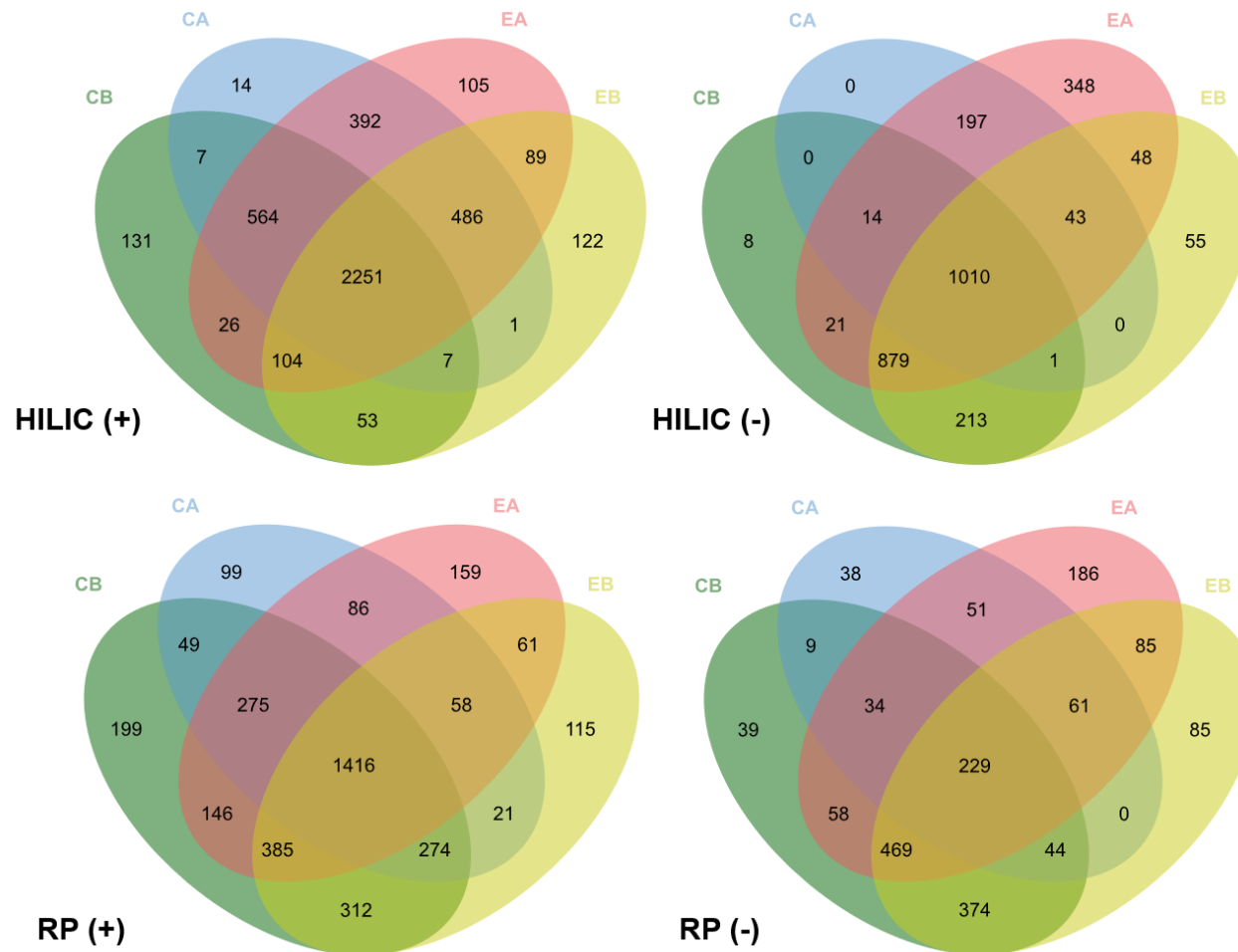
Figure 31: Venn diagrams showing overlapping HQFs between four cores for each LC/MS condition

C = *Carex*, E = *Eriophorum*, A = Site A, low-centered polygon, B = Site B, high-centered polygon

While the total number of features detected varied some between cores, overall, HILIC (+) detected the greatest number of HQFs across the four cores with 3929 (34.9 %) followed by RP (+) with 3618 (32.1 %), HILIC (-) with 2170 (19.3 %), and finally RP (-) with 1541 (13.7 %) (Figure 32). This is likely due to the more favorable ionization conditions in positive-mode, and more reproducible retention on the HILIC columns for the small, highly-polar compounds that dominate LMW DOM.[185] Despite some differences in performance, the optimized LC/MS conditions were still highly complementary with just 94 (2 %) HQFs observed by all four conditions (Figure 33). There was more overlap between LC phases within the same polarity—22.7 % overlap between HILIC and RP in positive-ion mode and 19.5 % overlap for negative-mode—than for the opposite polarities on the same LC phase—10.7 % overlap between positive- and negative-mode on the HILIC columns, and 7.1 % for the RP columns. Taken together, these results confirm that the dual-LC, dual-polarity approach is effective at expanding coverage of the LMW DOM pool and is sensitive enough to capture both shared features as well as those unique to Arctic soils obtained from different sampling sites with varying aboveground characteristics.

To broadly examine the variability across each of the LC/MS conditions prior to data filtering and normalization and evaluate the reproducibility of the untargeted approach across biological replicates, we built a correlation matrix using the calculated Pearson coefficients for each extract (Figure 34) and PCA plots using the unique identifiers and peak areas for each HQF (Figure 35). These are useful ways to visualize high-level similarities and differences (both qualitatively and quantitatively) across the dataset and
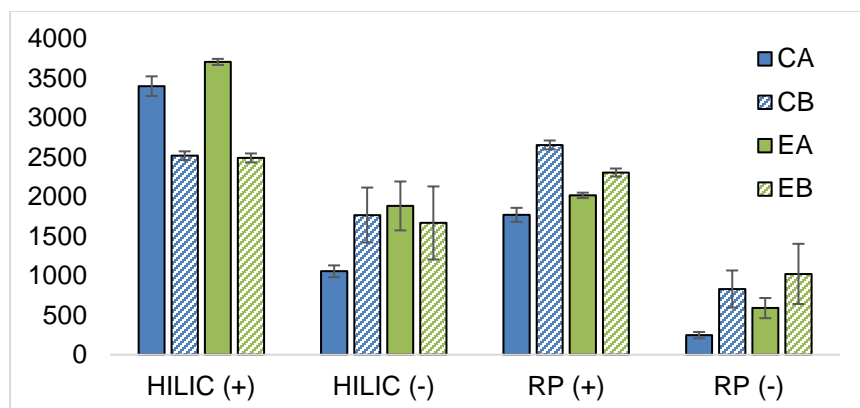
Figure 32: Number of HQFs observed in each core separated by LC/MS condition

C = *Carex*, E = *Eriophorum*, A = LCP, B = HCP
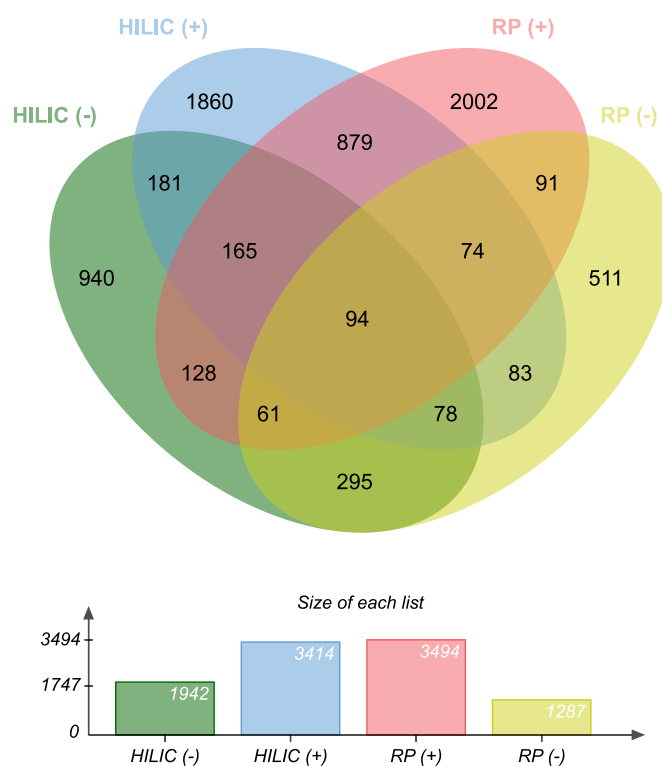
Figure 33: Venn diagram (top) showing overlap of HQFs between LC/MS conditions across all four cores and bar graph (bottom) showing total number of unique HQFs observed by each LC/MS condition
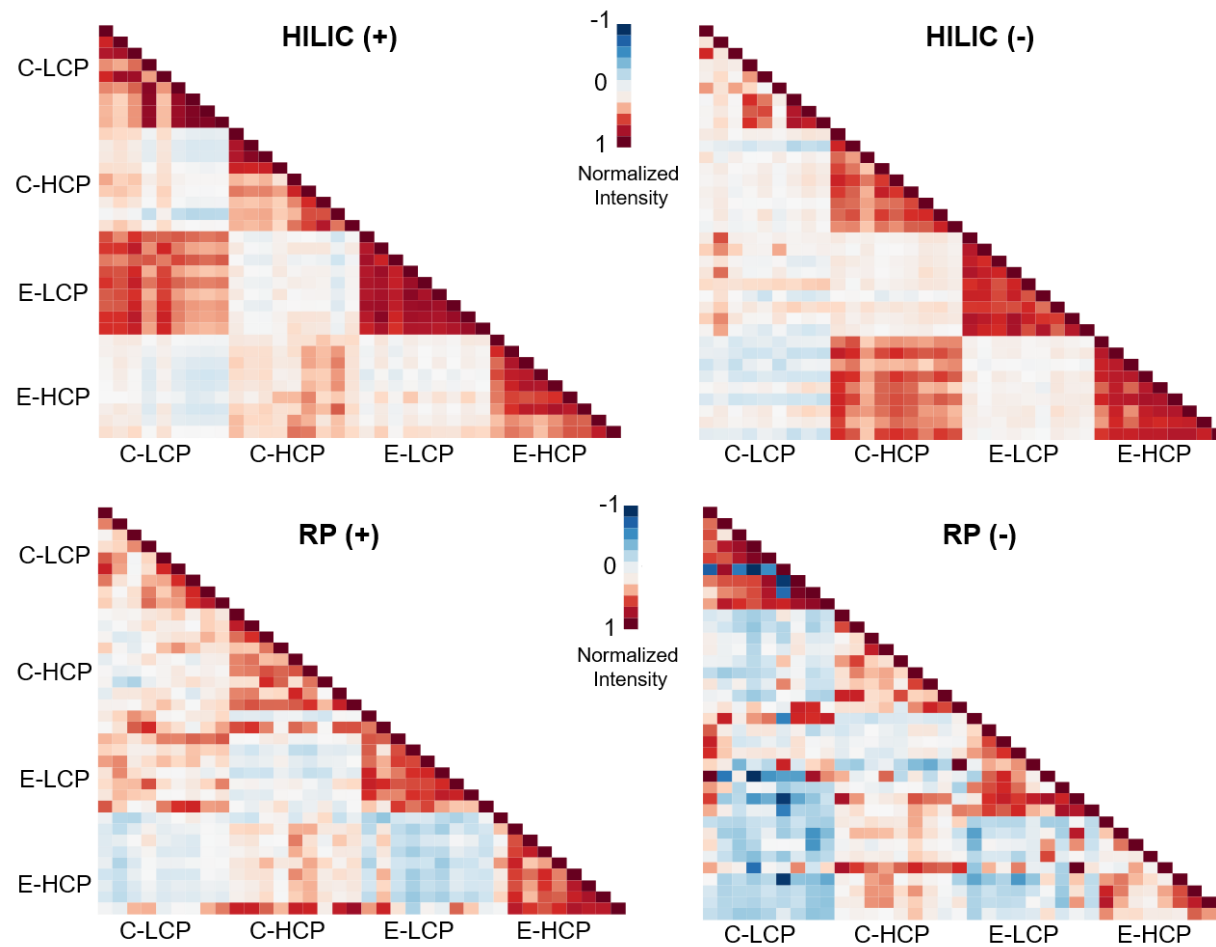
Figure 34: Pearson correlation plots of normalized $\log_2$ peak areas for the 36 samples analyzed, separated by each LC/MS condition
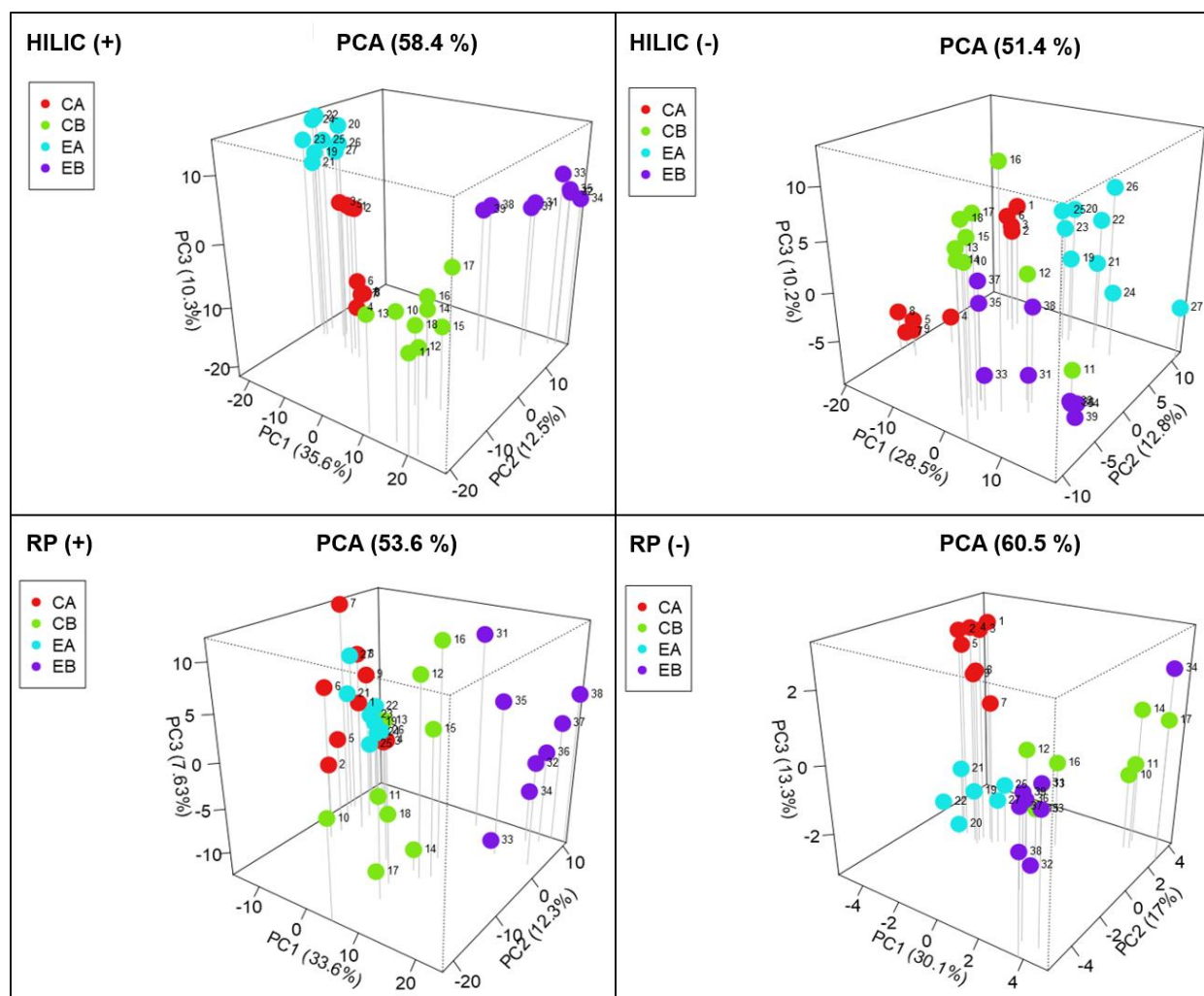
Figure 35: PCA of HQFs detected in each core separated by LC/MS condition

see how they relate to each other across the dataset. While there was some variability among replicates, which was more noticeable in the RP datasets, in general, there was a fair amount of correlation across each of the nine samples within each core (Figure 34). Interestingly, while one may expect that aboveground vegetation dictates belowground SOM composition, for all four LC/MS conditions, the cores from the same polygon type were more highly correlated to one another than the cores with the same aboveground vegetation, suggesting polygon type may be a stronger predictor of LMW DOM availability than vegetation cover at this scale (Figure 34). This also indicates that LMW DOM may be dominated by soil-derived organic species instead of plant inputs at these locations. Similarly, while there were various areas of overlap, visualization by PCA for each LC/MS condition generally revealed four identifiable clusters corresponding to each core (*Carex* – LCP, *Eriophorum* – LCP, *Carex* – HCP, and *Eriophorum* – HCP) suggesting unique LMW DOM profiles at each site (Figure 35). The components accounted for 58 %, 51 %, 54 %, and 61 % of the variation across the datasets for HILIC (+), HILIC (-), RP (+), and RP (-), respectively, indicating both polygon type and vegetation have a major effect on the LMW DOM composition. Like the correlation matrix though, cores from the same polygon type clustered closer together than those with the same vegetation, further supporting polygon type as a stronger predictor of LMW DOM composition and that it is a useful scaling parameter to connect biogeochemical measurements with landscape properties (*i.e.* thaw depth, hydrology).[50, 192, 195] The plots for both of the HILIC datasets also revealed some variation with depth indicated by two clusters of the extracts in red

109

(Figure 35). These results suggest that LMW DOM composition is influenced at a finer scale by depth as well.

4.4.2   *Impacts of polygon type and vegetation on LMW DOM availability*

To reduce the dimensionality of these data and identify features that were significantly differentially-abundant between cores, we performed pairwise comparisons by t-test and fold change analysis between cores of the same polygon or vegetation type, followed by an ANOVA to determine features that were in higher relative-abundance uniquely due to polygon type or vegetation (p-value < 0.001, FC > 4). To visualize these differentially-abundant features, we first used volcano plots to isolate the features that had the greatest FC and lowest p-value between conditions (Figure 36). There were more features found in higher relative abundance in the *Eriophorum* cores versus the *Carex* cores and at the HCP sites versus the LCP sites (Figure 36). The lower abundance at the LCP site may be due to increased transport (horizontal or vertical) of LMW DOM out of the organic horizon,[175, 199] likely due to the lower topography and more saturated conditions (Figure 37) or increased microbial processing. The higher relative abundance of LMW DOM features in the *Eriophorum* cores suggests either an accumulation or increased availability of LMW DOM, possibly due to higher dry root weight (Figure 37), which has been shown to enhance substrate availability.[200] Alternatively, this could be viewed as a depletion of LMW DOM in the *Carex* cores, which could be due to increased microbial processing, or plant uptake, of DOM at those sites. Plant uptake of DOM has been observed in Arctic vegetation before as a way for plants to overcome nitrogen limitation,[27, 28, 201] however, the total nitrogen (TN) measurements collected here were not significantly different between

110
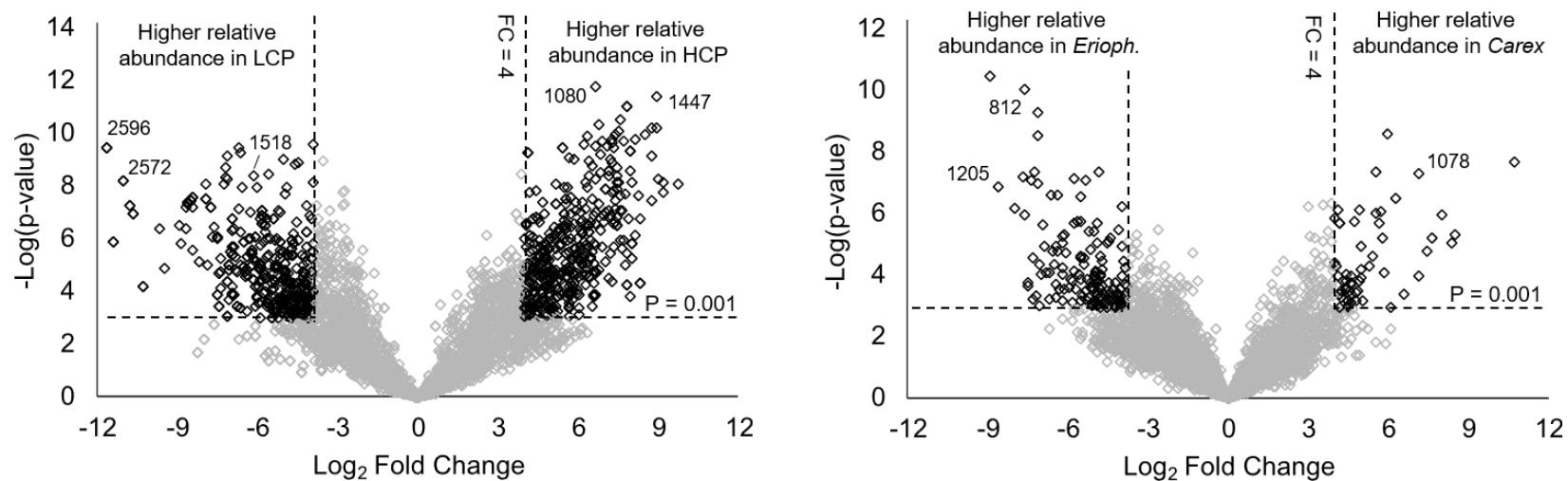
Figure 36: Volcano plots showing differentially-abundant LMW DOM features due to polygon (left) or due to vegetation (right) highlighting features that had a FC > 4 and passed the paired t-test p-value < 0.001 (dotted lines)

Inset numbers indicate unique identifiers called out in Table 10 below with molecular and annotation data.

111
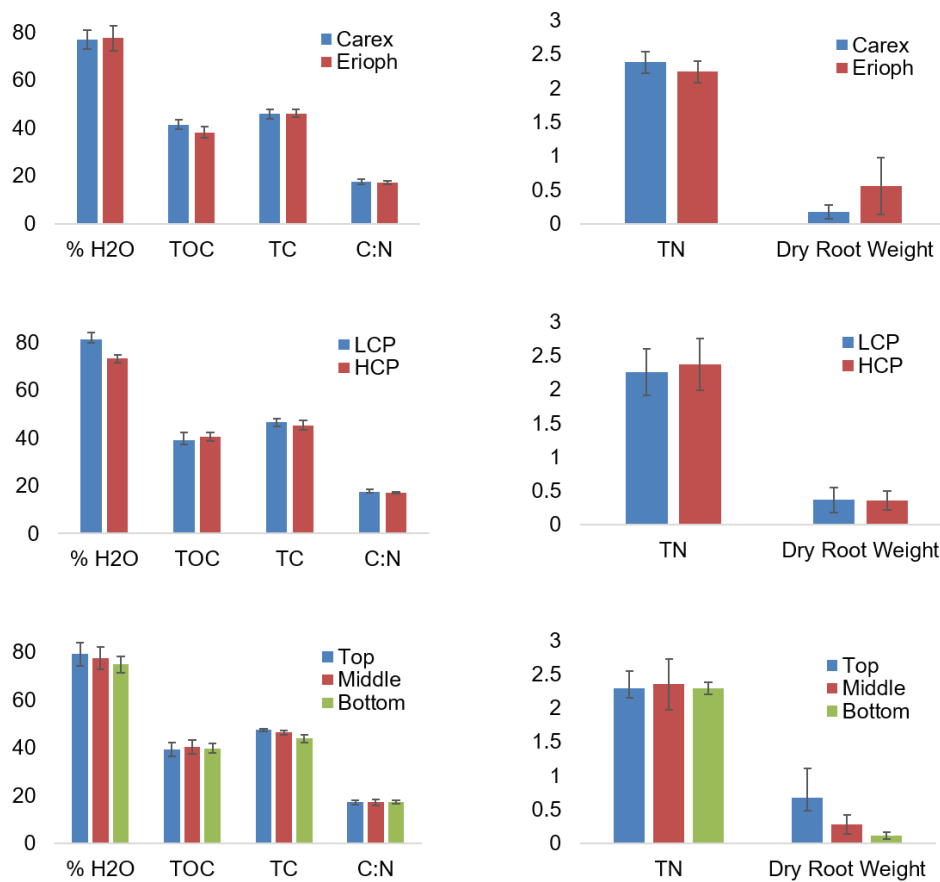
Figure 37: Visual summary of data from Table 7, including variation in % H_2O, TOC, TN, TC, C:N, and dry root weight between cores

Note: Y-axis is unitless because different units were used for different measurements (as shown in Table 7)

the *Carex* and *Eriophorum* cores (Figure 37), suggesting DON uptake was not a significant driver of LMW DOM variation at the time of collection. Further analysis of the molecular details (see next section) or additional studies that include gas flux measurements could be conducted to verify increased mineralization of LMW DOM at the LCP and *Carex* sites.

To evaluate the quantitative reproducibility of the differentially-abundant features, an analysis of the coefficient of variance (CV %) for the peak areas across replicates revealed that 95 % of these differentially-abundant features showed acceptable reproducibility (CV < 10 %, Figure 38) indicating the optimized data collection and processing techniques were robust and that the data filtering protocols were conservative, selecting for LMW DOM features that were consistently detected across replicates. It is important to note that some variability observed among replicates is not unexpected. Despite the subsamples of soil being relatively small (4 g), it has been well-established that LMW DOM composition and abundance can vary at even the micro-site or aggregate scale (10s-100s if µm).[140, 202] That the untargeted approach applied here can detect these subtle differences is an added benefit, as it demonstrates the sensitivity of the technique to detecting variation in the availability of LMW DOM across space and capturing both the biotic and abiotic impacts on this pool.

4.4.3   *Molecular characterization of differentially-abundant LMW DOM features*

We further investigated the relationship between polygon type or vegetation and LMW DOM availability by directly contrasting the differentially-abundant LMW DOM features using molecular data obtained from the high-resolution LC/MS measurements. Differentially-abundant features ranged in molecular weight (~56 – 900 *m/z*) and polarity, exhibited by their elution across the full retention time window for each LC/MS condition (Figure 39). However, the *m/z* distribution did not vary appreciably between cores or between the operationally-defined depth increments we employed (Figure 40). While these data support that the LC/MS conditions were not biased toward any particular class of com-
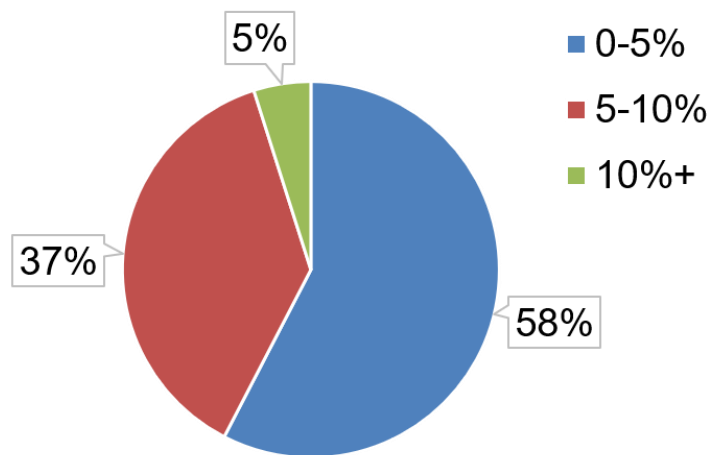
Figure 38: Proportion of differentially-abundant HQFs that had a CV between 0 – 5 % (blue), 5 – 10 % (red), or over 10 % (green)
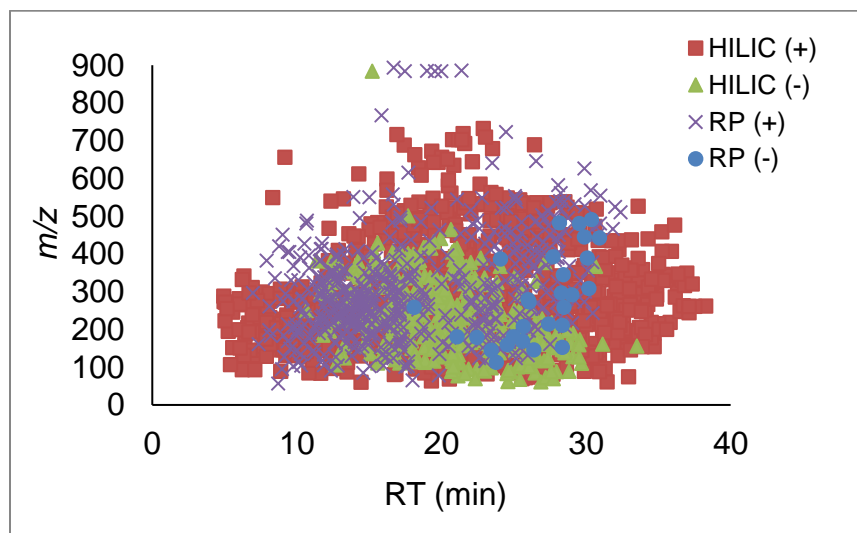
Figure 39: Distribution of MW and RT for differentially-abundant features due to polygon or vegetation, detected across all 36 extracts, separated by LC/MS condition
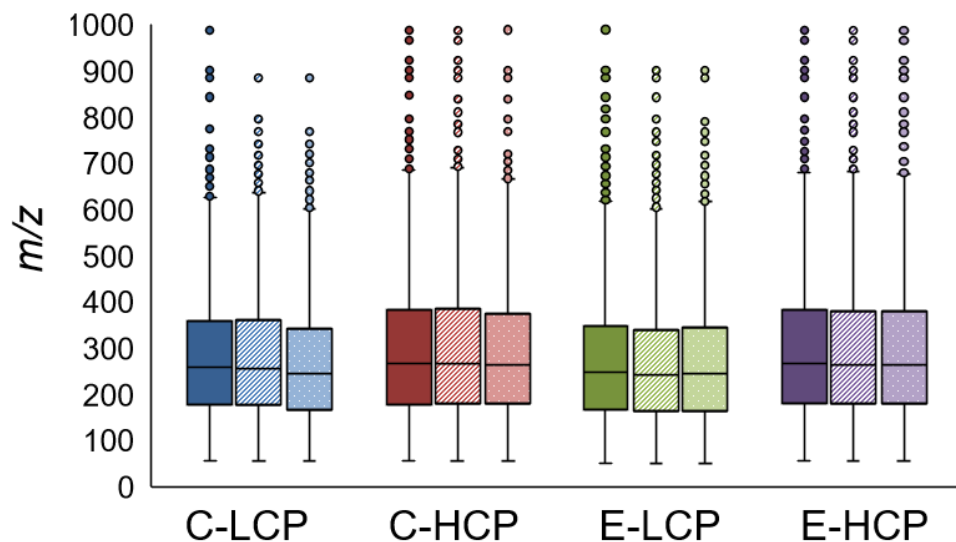
Figure 40: Distribution of *m/z*'s of differentially-abundant HQFs by core and depth
From left to right, solid color indicates "top", stripes indicate "middle", and dots indicate
"bottom" of the organic horizon; blue = *Carex* LCP, red = *Carex* HCP, green = *Erioph*
LCP, and purple = *Erioph* HCP

-pounds, they also indicate that molecular weight alone is not adequate at describing LMW
DOM availability across space and that additional molecular information is required.

Accordingly, using the high mass accuracy MS[1] measurements (< 5 ppm), we
assigned a molecular formula to any differentially-abundant feature meeting the criteria
outlined in Chapter 2, using C, H, N, O, P, and S, and calculated the double bond-
equivalents (DBE), aromaticity index (AI), and elemental ratios (H/C, O/C, N/C, and O/S).
The equations used to calculate DBE and AI are shown below in Equations 5 and 6:

$$\text{Equation 5:} \quad DBE = 1 + C - 0.5H + 0.5N + 0.5P$$

$$\text{Equation 6:} \quad AI = \frac{1+C-0.5O-S-0.5H}{C-0.5O-S-N-P}$$

Of the 521 differentially-abundant features, 217 (42 %) were assigned molecular formulas while 304 (58 %) did not meet the criteria for a confident assignment or were possible adducts, complexes, or fragments identified by the MZmine modules during annotation. As described above, approximately 10 % may have been adducts; including sodium (Na^+) or chloride (Cl^-) adducts as these are commonly seen in the characterization of OM using positive- and negative-ESI, respectively.[203, 204] Alternatively, these soils have also been shown to have high iron concentrations,[66, 205, 206] and since organo-iron complexes can be soluble in soil water, they may have been extracted here as part of the LMW DOM pool. Because organo-metal complexes generally dissociate upon ionization however, they would not appear in the mass spectrum, or would appear as an ion ([M-Fe+H]$^+$) less the mass of iron (55.9349 *m/z*) requiring an additional calculation and search to annotate these. Across the 217 assigned features, the average mass error was just 0.65 ppm and the average molecular weight was 379.9353 *m/z* (Table 10).

Elemental data were then used to assign a biomolecular compound class to each differentially-abundant feature based on their H/C and O/C ratios —lipids, proteins (amino acids and amino sugars), lignins, carbohydrates, unsaturated hydrocarbons, condensed aromatics (phenolics), tannins, and aliphatics. Using these data, a van Krevelen plot was built to help visualize the distribution of these classes across the four cores (Figure 41).[81, 207] Upon visual inspection of the plot, there is a clear density of formulas in the low O/C and high H/C regions of the plot, indicating an abundance of aliphatic compounds, such as
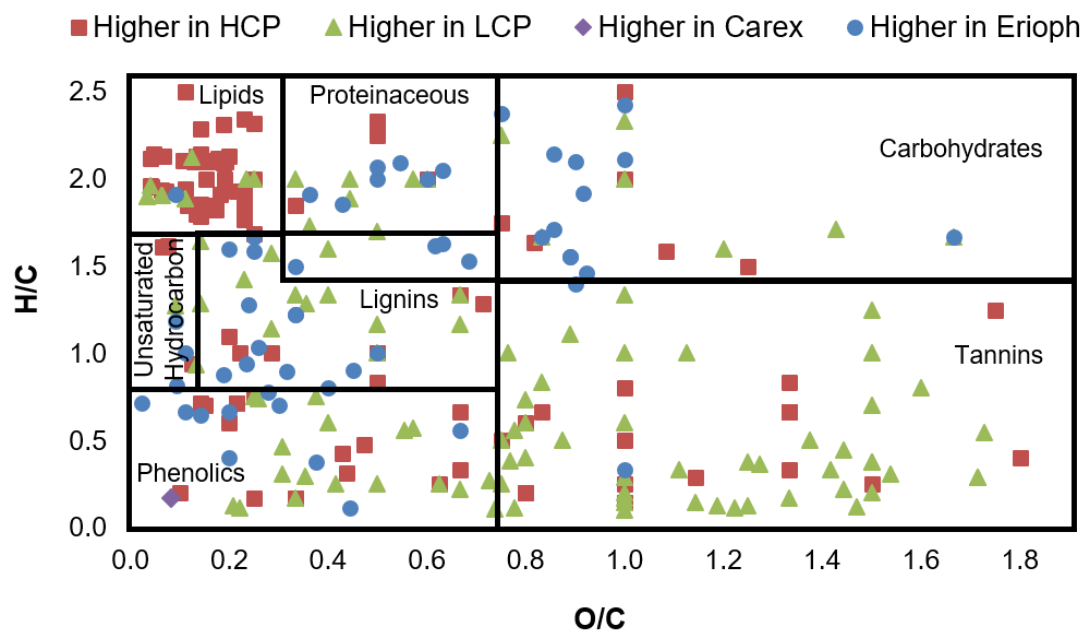
Figure 41: van Krevelen plot for molecular formulas assigned to differentially-abundant HQFs due to polygon type or vegetation

Boxes overlaid on the plot indicate assigned biochemical classes (based on Ohno *et al.* 2014 and Antony *et al.* 2014):[135, 208] lipids (O/C < 0.3, H/C > 1.7), peptides, amino acids, and amino sugars (0.3 < O/C < 0.7, H/C > 1.5), carbohydrates (O/C > 0.7, H/C > 1.5), unsaturated hydrocarbons (O/C < 0.1, 0.7 < H/C < 1.7), lignins (0.1 < O/C < 0.7, 0.7 < H/C < 1.7), tannins (O/C > 0.7, H/C < 1.5), and phenolics/condensed aromatics (O/C < 0.7, H/C < 0.7).

lipids, sugars, and amino acids possibly derived from microbial biomass. The high presence of formulas consistent with phenolics, lignins, and proteinaceous (*i.e.* peptides, amino sugars) material is indicative of freshly-deposited plant material. As both *Carex* and *Eriophorum* are vascular plant species and decomposition is generally slowed in Arctic systems, an accumulation of lignified LMW DOM across cores was anticipated. While the differences between the *Carex* and *Eriophorum* cores were difficult to compare due to most

117

of the differentially-abundant features being more abundant in the *Eriophorum* cores, in general, the compounds that were found in higher relative abundance were dominated by formulas consistent with low O/C and high H/C (*i.e.* aliphatic) content as well. This may have been due to the higher root biomass or, more likely, necromass found beneath the *Eriophorum* cores (Figure 37), which may have led to higher aliphatic content due to higher root exudation or the buildup of common root tissue components upon root death.[209]

While the H/C vs O/C van Krevelen plot was used here as a high-level approach to visualize variation in the biomolecular classes of compounds present in these soils, methods have recently been proposed to improve biomolecular assignment of molecules from ecological samples, for example, by including N and P as well.[210] Since N-containing compounds made up over 70 % of the differentially-abundant features detected at each polygon and are the most vulnerable to microbial degradation, here, we have also included a van Krevelen analysis between the two polygon types using the N/C ratio (Figure 42); although, this technique may also be used with the S/C or P/C ratios to visualize the distribution of heteroatoms across LMW DOM features detected. Using this approach, the results show a clear separation between the N-containing features at the LCP and HCP sites. More features with a low N/C ratio (N/C < 0.2) and high H/C content (H/C > 1.5), which are consistent with lipid-like compounds, were found at the HCP site, consistent with our findings from above. Features in the region N/C < 0.1 and H/C < 1.5, indicating a high number of amino groups and the presence of phytochemicals (bioactive plant compounds),[210] were similar between the two polygon types. Features with higher N/C > 0.2, consistent with LMW DOM compounds having secondary or tertiary amines (*i.e.*
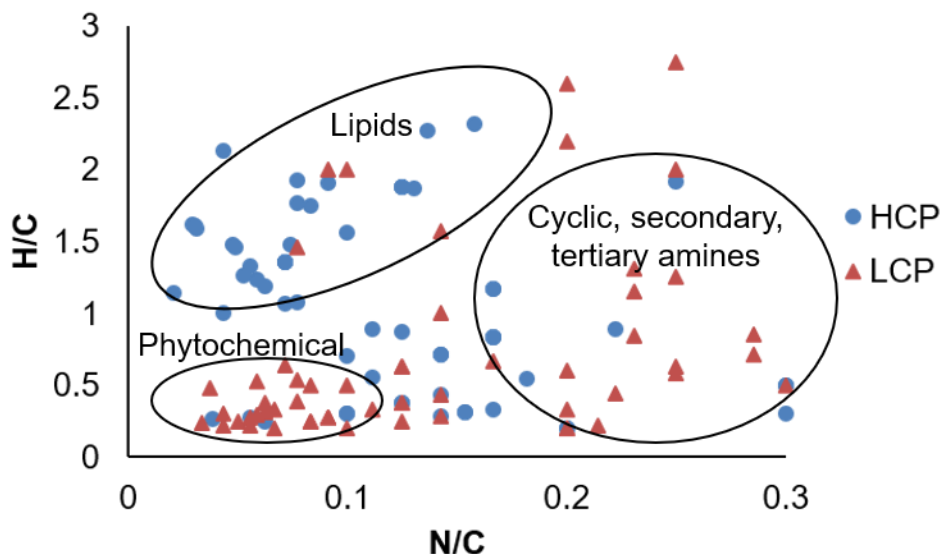
Figure 42: van Krevelen plot using N/C ratio instead of O/C ratio to explore nitrogen dynamics in Arctic LMW DOM extracts

alkaloids, cyclic amines), were more dominant at the LCP site, consistent with our hypothesis from Chapter 3 that the *Carex* core may have experienced alkaloid stress.

For a more detailed view of the LMW DOM chemistry at these sites, the average molecular properties for the differentially-abundant features that were assigned formulas have been reported (Table 10). Due to polygon type being a stronger predictor of LMW DOM availability, features that were in higher relative abundance at either the HCP or LCP sites have been highlighted. In contrasting the two polygon types, there were readily-observable differences reflected in the LMW DOM pool. Consistent with our hypothesis above from the van Krevelen analysis, there are multiple lines of evidence to support increased microbial processing and C cycling at the LCP site. First, both the average *m/z* and DBE were lower in the LCP cores, characteristic of SOM that has undergone microbial

Table 10: Average molecular properties for HQFs that were in higher relative abundance due to polygon type or vegetation Average formula, m/z, DBE, AI, element heuristics, and proportion of biochemical classes determined using high-resolution mass spectral data for LMW DOM features that were consistently and significantly more abundant in the HCP and LCP are reported in addition to molecular data for the differentially-abundant features due to vegetation and across all assigned features

	HCP	LCP	Due to Vegetation	All differentially-abundant features
Number of features	92	95	30	217
Formula	$C_{17}H_{21.5}O_{4.9}N_{1.2}S_{2.0}P_{0.2}$	$C_{11.9}H_{7.8}O_{6.5}N_{1.1}S_{2.3}P_{0.5}$	$C_{20.3}H_{20.6}O_6N_{0.9}S_{0.5}P_{0.2}$	$C_{15.2}H_{15.4}O_{5.7}N_{1.1}S_{1.9}P_{0.3}$
m/z	393.0249	361.4926	398.1955	379.9353
DBE	7.88	9.51	11.5	9.09
AI	0.37	0.75	0.42	0.54
H/C	1.12	0.75	1.12	0.81
O/C	0.45	3.40	0.38	0.62
N/C	0.15	0.17	0.08	0.20
O/S	2.6	3.4	5.13	3.26
DBE/C	0.59	0.79	0.56	0.67
DBE/H	1.15	2.10	0.76	1.51
DBE/O	3.2	2.35	3.62	2.89
C:N	14.4	10.9	23.4	13.9
% lipid	16.3	2.11	6.67	8.75
% protein	7.61	6.32	6.67	6.91
% lignin	13.0	4.21	30.0	11.5
% carbohydrate	5.43	6.32	6.67	5.99
% unsaturated	12.0	4.21	10.0	8.29
% aromatic	30.4	51.6	33.3	40.1
% tannin	15.2	25.3	6.67	18.4

decomposition. Second, the proportion of features characteristic of compounds with higher biodegradability—lipids, carbohydrates, aliphatics—were also lower in the LCP cores, suggesting they may have been preferentially-degraded and released as GHGs.[63, 64, 198] Third, there was a higher relative abundance of tannins and other condensed aromatics at the LCP site, as shown in Figure 41 and by the higher AI and 25.3 % tannin content shown in Table 10, suggesting an accumulation of these more recalcitrant features. Finally, although LCP centers are generally more anaerobic due to saturated conditions, the average oxygen content (demonstrated by the average molecular formula, O/C, and O/S ratios) of the differentially-abundant LMW DOM features at the LCP was higher than the HCP, further supporting enhanced microbial processing of OM at the LCP site. Taken together, these results reveal a detailed picture of C and N cycling at these sites, yielding insight into the chemical processing and relative degradability of the LMW DOM features found across the Arctic landscape.

A selection of LMW DOM features that had the highest fold change between sites have been summarized in Table 11 below. Interestingly, of the assigned formulas at the HCP and LCP sites, 88 % and 72 %, respectively, contained N, suggesting root exudation of organic N may be an important process occurring at these sites, especially in the HCP cores. This could be a result of the *priming effect* discussed in Chapter 1. Because HCP polygons are drier, plant and microbial activity may be more limited. As such, vegetation may allocate more N belowground to try and stimulate microbial processing of organic matter to release nutrients for uptake.[76, 211] Somewhat surprisingly, ~11 % of the differentially-abundant formulas across all four cores contained both sulfate and nitrate

groups (O > 6), which are characteristic of secondary organic aerosols (SOAs).[212] Secondary organic aerosols are formed in the atmosphere through a complex interaction of sunlight and volatile organic compounds that originate from industrial emissions, cars, burning biomass, or even vegetation.[213] They have been shown to be an important input of organic C to alpine systems where they influence a range of biogeochemical processes.[214] However, while they have been observed near Utqiaġvik before, it has generally been along the coastline or in the marine environment closest to anthropogenic activities.[215] To the best of our knowledge, this is the first evidence of water-soluble SOAs in polygonal tundra soils on the BEO. These results suggest that some portion of LMW DOM that is available for microbial processing is derived from volatile organic carbon precursors.

Also of note, although there were a similar number of chemical formulas detected in higher relative abundance at each polygon, the features at the HCP site were more chemically diverse as indicated by a more equitable distribution among the assigned compound classes (Table 11). One explanation for this is that although the aboveground vegetation in each core represented primarily a single species, the HCPs generally have higher plant diversity. This has been associated with more diverse plant inputs into the soil and increased microbial diversity, in turn leading to a more diverse substrate pool.[216]

Another way to examine the differentially-abundant features is to distinguish clusters of features that vary similarly across cores using two-way hierarchical clustering with the normalized $\log_2$ peak areas and a unique identifier for each feature (Figure 43). This allows for a more detailed view of LMW DOM features that were consistently and

Table 11: A selection of LMW DOM features detected in higher relative abundance at each of the sites

The unique identifier, LC/MS condition by which the feature was detected, it's *m/z*, the predicted formula and class of compound along with the ppm error and fold change between cores are reported.

Site with higher relative abundance	Unique ID	LC/MS Condition	*m/z*	Predicted Formula	ppm error	Class	Fold Change (*Erioph* cores)	Fold Change (*Carex* cores)
LCP	2596	HILIC (-)	347.8884	$C_9H_2O_9P_1S_2$	-4.7	Tannin	-11.63	-7.645
LCP	2572	HILIC (-)	400.8711	$C_7H_2N_2O_{12}S_3$	3.5	Tannin	-11.01	-6.884
LCP	2599	HILIC (-)	367.8842	$C_8H_3NO_{10}S_3$	-1.1	Tannin	-10.61	-8.179
LCP	1429	HILIC (-)	400.8702	$C_7H_2N_2O_{12}S_3$	1.3	Tannin	-10.16	-6.893
LCP	1518*	HILIC (-)	192.0527	$C_7H_7N_5O_2$	0.1	Lignin	-	-5.111
HCP	3940	HILIC (+)	702.5350	$C_{50}H_{68}O_2$	0.2	Unsaturated Hydrocarbon	9.189	8.014
HCP	2122	HILIC (+)	506.8323	$C_{13}H_4N_2O_5S_7$	-1.9	Condensed Hydrocarbon	7.311	8.501
HCP	1690*	RP (+)	273.2535	$C_{15}H_{32}N_2O_2$	-0.7	Lipid	5.685	7.433
HCP	1080	RP (+)	453.3682	$C_{26}H_{48}N_2O_4$	-1.1	Lipid	6.616	9.623
HCP	1447*	RP (+)	104.0705	$C_4H_9NO_2$	-1.2	Protein	8.020	8.594

Site with higher relative abundance	Unique ID	LC/MS Condition	*m/z*	Predicted Formula	ppm error	Class	Fold Change (HCP cores)	Fold Change (LCP cores)
Erioph	771	HILIC (+)	286.1138	$C_9H_{19}NO_9$	1.8	Carbohydrate	-5.56161	-7.98873
Erioph	812*	HILIC (+)	363.0902	$C_{10}H_{14}N_6O_9$	1.9	Tannin	-	-7.38981
Erioph	790*	HILIC (+)	251.0764	$C_9H_{14}O_8$	0.9	Carbohydrate	-6.57539	-6.22072
Erioph	2435	RP (+)	459.1955	$C_{32}H_{26}O_3$	0	Unsaturated Hydrocarbon	-7.07477	-
Erioph	1181	RP (+)	548.2498	$C_{11}H_{21}N_{27}O$	0.5	Lipid	-7.52305	-
Erioph	1205	RP (+)	550.2340	$C_{18}H_{27}N_{15}O_6$	-0.3	Lignin	-8.61341	-
Erioph	1267	RP (+)	226.1285	$C_8H_{19}NO_6$	-0.2	Carbohydrate	-7.70772	-
Carex	1078	HILIC (+)	191.0233	$C_{12}H_2N_2O$	-3.8	Aromatic	7.1385	-
Carex	2103	HILIC (-)	236.8647	no hit	-	-	4.19787	-
Carex	2500	HILIC (-)	414.7706	no hit	-	-	4.12659	-

*Indicates feature has been annotated by database searching (see Appendix C)
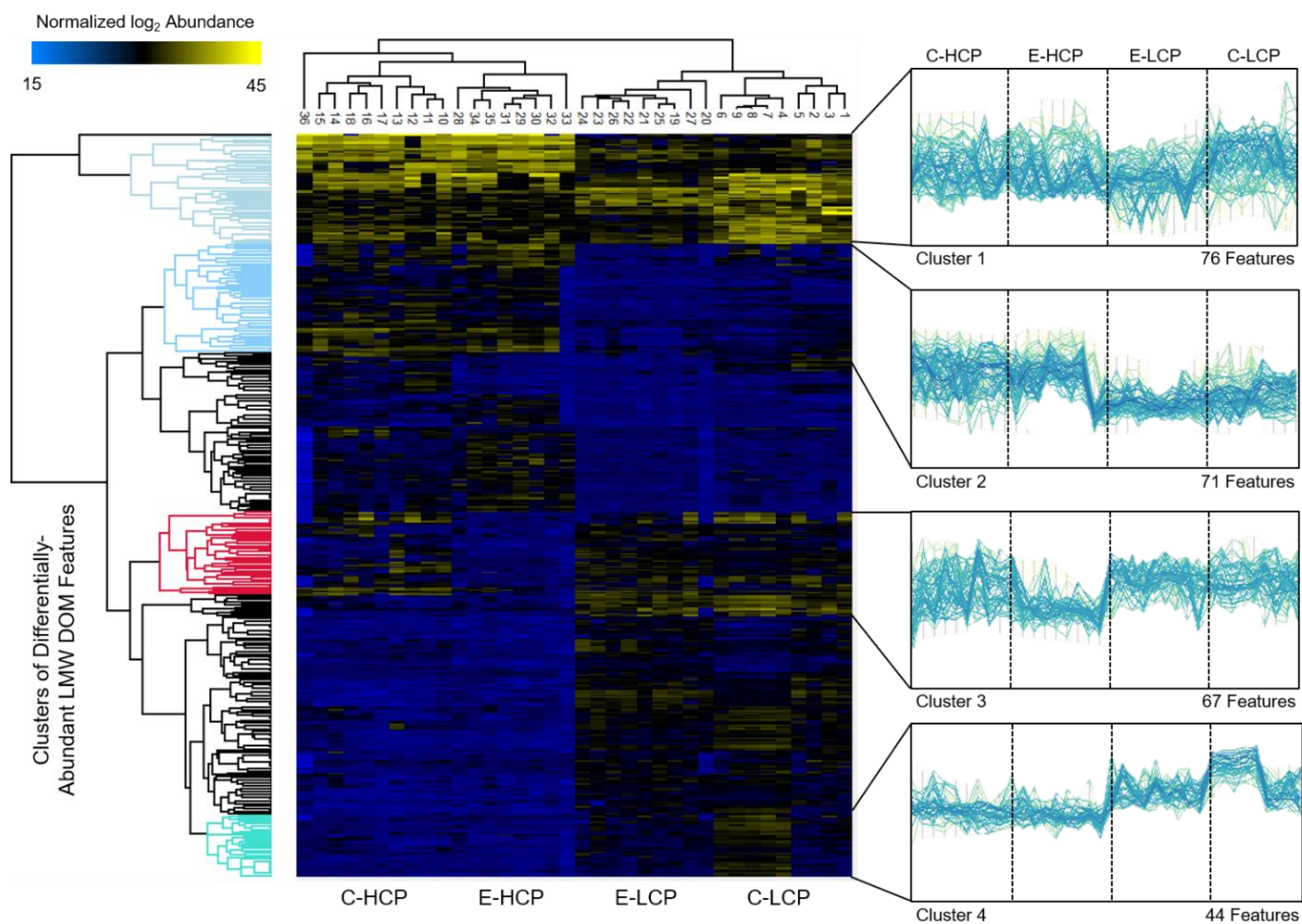
Figure 43: Two-way hierarchically-clustered heat map of normalized $\log_2$ abundances for 521 differentially-abundant LMW DOM features; four clusters have been called out to the right showing four main trends in the data

similarly varying across space. Consistent with previous analyses described above, the cores clustered into two main groupings corresponding to the cores from the same polygon type, LCP or HCP, samples 1-9 with 19-27 and samples 10-18 with 28-36, respectively. The LMW DOM features also clustered into groups based on their relative abundance variations across the cores. Four clusters have been highlighted to show the subtle, but consistent and significantly-different variations between cores due to polygon type, vegetation, or in some cases, depth (Figure 43). For example, cluster 1 shows 76 features that are somewhat abundant across most of the cores except for the *Eriophorum* core at the LCP site where those features were found in lower relative abundance. Cluster 2 shows 71 features that were depleted in both LCP cores but not the HCP cores. Cluster 3 indicates that 67 features were depleted in the *Eriophorum* core at the HCP site, and cluster 4 shows 44 features that were in higher relative abundance in the LCP cores, but that this varied with depth in the *Carex* core at that site.

In cluster 1, of the 76 differentially-abundant features, 49 (64 %) were assigned a chemical formula (average mass error = 0.406 ppm) based on high mass accuracy MS[1] measurements, and 4 others that were not assigned a chemical formula but did match to a database (< 5 ppm), for a total of 53 (70 %) features annotated in the cluster (Table 12). Among the LMW DOM features annotated in this cluster by database matching, there were amino acids, plant hormones, microbial metabolites, lignin-like molecules, and DNA/RNA fragments/derivatives. Importantly, these data support that this approach can detect key compounds involved in biogeochemical cycling. For example, a urea derivative (N-hydroxymethyl urea, [M-H]⁻ detected at 89.0358 *m/z*), was found to be in higher relative

125

Table 12: Cluster of annotated, differentially-abundant LMW DOM features found in high relative abundance in every core except the *Eriophorum* – HCP core

m/z	Predicted Formula	Mass error (ppm)	Compound Class	Database Annotation	Database Formula	Database Compound Class	MW
321.0933	$C_{11}H_{18}N_2O_9$	-2	carbohydrate	-	-	-	-
275.9782	$C_{11}H_3NO_8$	-1.3	tannin	-	-	-	-
247.9740	$C_{11}H_6N_4O_{19}$	0	tannin	-	-	-	-
325.1183	$C_{12}H_{19}N_6O_3P$	-0.1	lignin	-	-	-	-
191.5355	$C_{12}H_{19}NO_{13}$	-0.1	carbohydrate	-	-	-	-
380.0831	$C_{13}H_{19}NO_{12}$	-0.6	tannin	-	-	-	-
265.0606	$C_{15}H_{10}N_2O_3$	-3.5	aromatic	6-acetophenazine-1-carboxylic acid	$C_{15}H_{10}N_2O_3$	aromatic	266.069
281.0920	$C_{16}H_{14}N_2O_3$	-3	lignin	-	-	-	-
311.1029	$C_{17}H_{16}N_2O_4$	-1.6	lignin	-	-	-	-
110.9824	$C_{17}H_{36}N_2O_3$	0.7	lipid	-	-	-	-
511.4389	$C_{17}H_{50}N_{16}O_2$	2.6	lipid, aliphatic	-	-	-	-
337.0826	$C_{18}H_{14}N_2O_5$	0.1	lignin	-	-	-	-
145.0889	$C_{18}H_{27}N_{15}O_6$	-0.3	lignin	-	-	-	-
225.9641	$C_{19}H_2O_{14}$	-3.6	tannin	-	-	-	-
226.9553	$C_2H_4N_4O_5S_2$	1.3	carbohydrate	-	-	-	-
272.9246	$C_2H_4N_4O_6P_2S$	-2.8	carbohydrate	-	-	-	-
214.8694	$C_2HO_4PS_3$	-3.5	tannin	-	-	-	-
267.0913	$C_3H_{12}N_{10}O_5$	-0.8	carbohydrate, aliphatic	-	-	-	-
253.1114	$C_3H_{14}N_{10}O_4$	-1.8	carbohydrate, aliphatic	-	-	-	-
223.9617	$C_3H_3N_3O_7S$	-0.8	tannin	-	-	-	-
246.9834	$C_3H_4N_8O_2S_2$	3.4	lignin	-	-	-	-
243.0215	$C_3H_8N_4O_9$	0.3	carbohydrate, aliphatic	-	-	-	-
268.9273	$C_4H_3N_2O_8PS$	-0.6	tannin	-	-	-	-

Table 12 continued

m/z	Predicted Formula	Mass error (ppm)	Compound Class	Database Annotation	Database Formula	Database Compound Class	MW
231.9430	$C_4H_3N_5OS_3$	1.4	lignin	-	-	-	-
145.0621	$C_5H_{10}N_2O_3$	-3.7	protein	alanine-glycine	$C_5H_{10}N_2O_3$	protein	146.069
117.0561	$C_5H_{10}O_3$	3.4	protein	2-hydroxy-3-methyl butyric acid	$C_5H_{10}O_3$	metabolite	118.063
240.0767	$C_5H_{15}N_5O_4S$	-3.6	carbohydrate, aliphatic	(3Z)-3-(1H-imidizol-5-ylmethylene)-5-methoxy-1H-indol-2(3H)-one	$C_{13}H_{11}N_3O_2$*	aromatic, protein	241.085
206.9522	$C_5H_5O_5PS$	-0.1	tannin	-	-	-	-
279.0931	$C_5H_8N_{14}O$	2.3	lignin	-	-	-	-
206.0707	$C_6H_{14}N_3O_3P$	3.5	protein	-	-	-	-
429.1308	$C_6H_{18}N_{14}O_9$	0.7	carbohydrate, aliphatic	-	-	-	-
415.1512	$C_6H_{20}N_{14}O_8$	0.4	carbohydrate, aliphatic	-	-	-	-
231.9466	$C_6H_4NO_5PS$	-3.8	tannin	-	-	-	-
211.0028	$C_6H_5N_4O_3P$	0.8	lignin	-	-	-	-
204.9729	$C_6H_7O_4PS$	-0.3	lignin	-	-	-	-
350.8737	$C_6HN_4O_6PS_3$	4	tannin	-	-	-	-
191.0535	$C_7H_{13}NO_3S$	-4.3	protein	-	-	-	-
253.0968	$C_7H_{18}N_4O_4S$	-2.7	protein, aliphatic	-	-	-	-
400.8702	$C_7H_2N_2O_{12}S_3$	1.3	tannin	-	-	-	-
400.8711	$C_7H_2N_2O_{12}S_3$	3.5	tannin	-	-	-	-
192.0527	$C_7H_7N_5O_2$	0.1	lignin	glucuronamide	$C_6H_{11}NO_6$*	carbohydrate	193.059
347.8884	$C_7HN_3O_8P_2S$	-0.7	tannin	-	-	-	-
367.8842	$C_8H_3NO_{10}S_3$	-1.1	tannin	-	-	-	-
416.8452	$C_8H_3O_{12}PS_3$	0.2	tannin	-	-	-	-

127

Table 12 continued

m/z	Predicted Formula	Mass error (ppm)	Compound Class	Database Annotation	Database Formula	Database Compound Class	MW
416.8453	$C_8H_3O_{12}PS_3$	0.4	tannin	-	-	-	-
180.0653	$C_9H_{11}NO_3$	-3.8	lignin	-	-	-	-
440.8636	$C_9H_2N_2O_{13}S_3$	-2.3	tannin	-	-	-	-
76.0592	$C_9H_9N_2O_2$	0.6	lignin	4-ethoxy carbonyl benzenediazonium	$C_9H_9N_2O_2$	aromatic	177.066
89.0358	no hit	-	-	N-(hydroxy methyl)urea	$C_2H_6N_2O_2$	metabolite	90.0429
128.0724	no hit	-	-	6-carboxypiperdine	$C_6H_{11}NO_2$	protein	129.079
138.0572	no hit	-	-	3-amino-2,3-dihydro benzoic acid	$C_7H_9NO_2$	protein	139.063
218.1063	no hit	-	-	(2Z)-2-methyl-4-(9H-purine-6-ylamino)-2-buten-1-ol	$C_{10}H_{13}N_5O$	plant hormone	219.112
94.9664	no hit	-	-	-	-	-	-
94.9666	no hit	-	-	-	-	-	-
102.0569	no hit	-	-	-	-	-	-
103.0540	no hit	-	-	-	-	-	-
110.9594	no hit	-	-	-	-	-	-
110.9765	no hit	-	-	-	-	-	-
112.0741	no hit	-	-	-	-	-	-
119.9483	no hit	-	-	-	-	-	-
124.9858	no hit	-	-	-	-	-	-
127.0539	no hit	-	-	-	-	-	-
133.0065	no hit	-	-	-	-	-	-
134.9178	no hit	-	-	-	-	-	-
147.0638	no hit	-	-	-	-	-	-
148.0022	no hit	-	-	-	-	-	-
149.9970	no hit	-	-	-	-	-	-

Table 12 continued

m/z	Predicted Formula	Mass error (ppm)	Compound Class	Database Annotation	Database Formula	Database Compound Class	MW
176.0906	no hit	-	-	-	-	-	-
196.9030	no hit	-	-	-	-	-	-
216.9125	no hit	-	-	-	-	-	-
236.8647	no hit	-	-	-	-	-	-
294.8031	no hit	-	-	-	-	-	-
324.7715	no hit	-	-	-	-	-	-
416.7706	no hit	-	-	-	-	-	-
488.8209	no hit	-	-	-	-	-	-

abundance at the HCP site. As a key metabolite in N cycling (*i.e.* ornithine cycle), urea is produced/excreted when there is an accumulation of highly toxic ammonia. An accumulation of extracellular urea in these soils may suggest increased inorganic N availability. Although further examination of the relative quantitative trends of other compounds involved in the urea cycle detected here (*i.e.* glutamate, glutamine, arginine, citrulline) would provide additional insight, this example demonstrates the utility of this untargeted approach in elucidating ecologically-relevant molecular information to be used in mechanistic modeling. When a compound was annotated by both elemental formula assignment and database matching, most of the time the formulas matched. However, there were instances where different formulas were assigned to the same molecule, which occurred twice in this cluster as well, indicated by the asterisks is Table 12. In these cases, we were able to use MS^2 fragmentation data to match to available data or eliminate incorrect assignments, highlighting the value of MS^2 data in providing information about both known (already in a database) and unknown compounds (or adducts/complexes for that matter). As an example, in the case of the [M-H]$^-$ ion detected at 192.0527 *m/z*, characteristic neutral losses of formamide (-CH_3NO, 45.0214 Da) and multiple dehydrations (-H_2O, 18.0098 Da) were observed, indicating a structure consistent with glucuronamide, a monosaccharide derivative of beta-D-glucuronic acid, a common microbial metabolite involved in ascorbic acid synthesis (Figure 44).[217]

It is important to note that although no formula or database match was made for the compounds at the bottom of Table 12, each of those features was reproducibly and reliably detected, and were robustly and conservatively determined to be significantly differentially

130
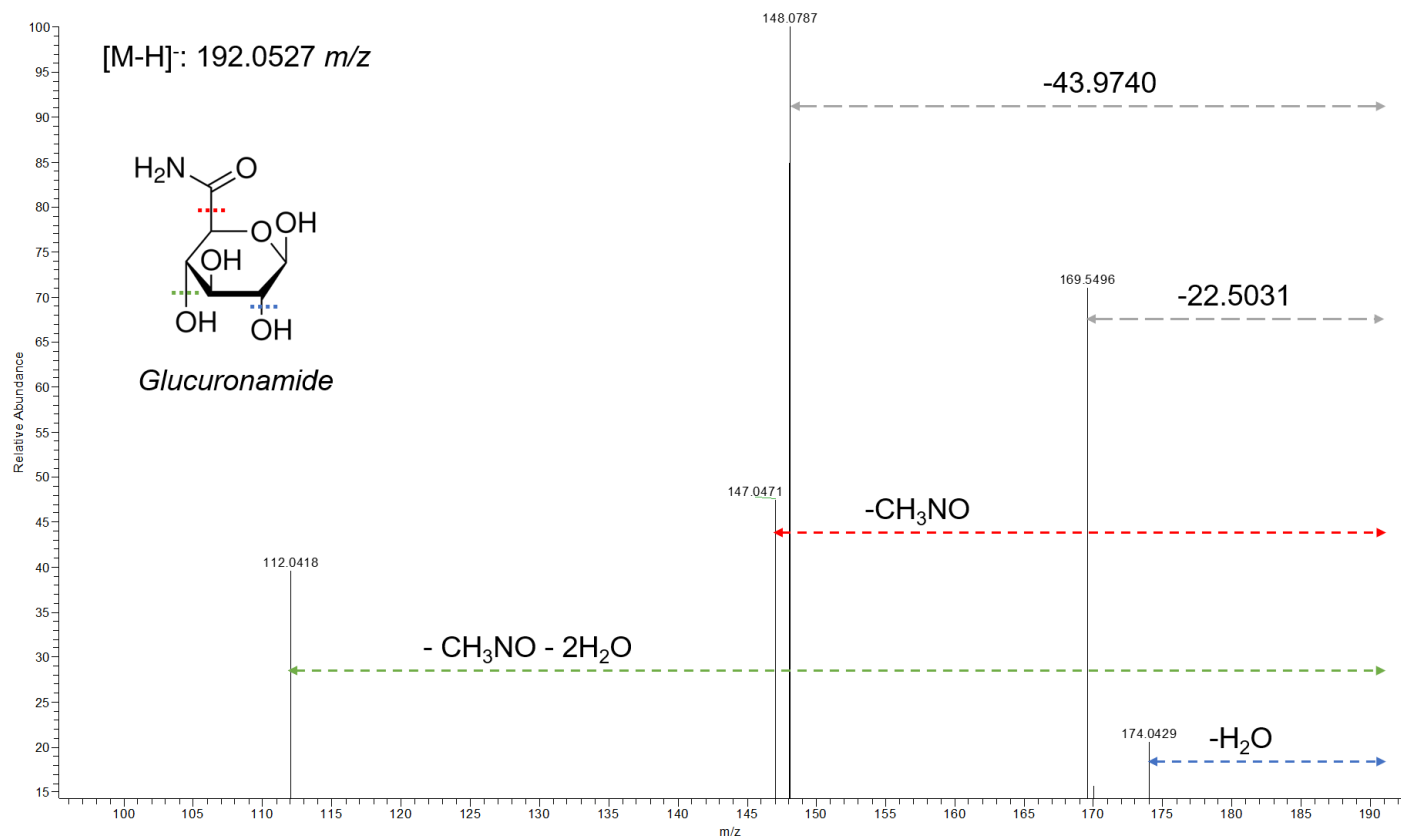
Figure 44: Fragmentation spectra of [M-H]⁻ ion at 192.0527 *m/z* showing characteristic neutral losses used for putative annotation

131

abundant between samples. In addition, each feature has a reproducible RT and peak area, and both MS^1 and MS^2 high-mass accuracy measurements. Thus, this is an information-rich signal that can be used diagnostically for both qualitative and quantitative research questions.

4.5   Conclusions

This study implemented the optimized dual-LC, dual-polarity LC/MS approach developed in Chapter 3 to examine the variation in LMW DOM availability in soil cores with two contrasting aboveground vegetation profiles and polygon types. These results support that a broad range of compounds with varying physicochemical properties and concentrations were detected by the optimized approach and that the untargeted platform is sensitive, robust, and reproducible even when applied across multiple cores from different sites across the landscape. We provide evidence that LMW DOM is a diverse and reactive pool, and while there were a common set of metabolites among the cores, there were significant differences observed between sites as well indicating LMW DOM may be an important driver of biogeochemical variation across the landscape. In addition, the untargeted LC/MS approach was sensitive to variation at multiple scales. While polygon type was a strong predictor of LMW DOM composition and availability, vegetation and depth also had an impact, indicating LMW DOM provides a window into the dynamic and complex interactions between landscape topography, vegetation, and SOM cycling.

Furthermore, this study revealed evidence of enhanced microbial processing at the LCP and *Carex* sites demonstrating its ability to detect hotspots of biogeochemical activity across space. Of the 521 differentially-abundant features detected, 217 were putatively

annotated by formula assignment, database matching, and evaluating the fragmentation data. For some compounds, this is the first time they have been reported in Arctic soils, including the 11 % of detected formulas consistent with secondary organic aerosols, although additional studies are needed to understand the relative importance of this process in these systems. With an average mass error of < 1 ppm, these high-mass accuracy measurements combined with reproducible retention times and peak areas provide an information-rich chemical profile of LMW DOM features in soil. Correlating these qualitative and quantitative variations with additional landscape-scale features (*i.e.* hydrology, gas fluxes) would yield additional insight into how this chemical signal may be used to predict various processes impacting C cycling in the Arctic.

CHAPTER 5: EVALUATING LMW DOM AVAILABILITY ACROSS AN ARCTIC PERMAFROST THAW GRADIENT

Portions of this chapter have been adapted from the following manuscript:

Ladd, M.P., Taş, N., Wullschleger, S.D. Hettich, R.L. Characterizing the vulnerability of low molecular weight dissolved organic carbon to release as greenhouse gases along an Arctic permafrost thaw gradient. *Soil Biology and Biochemistry* (*in prep*).

ML's contributions included: literature review, experimental design, sample collection, sample preparation, data collection, data analysis, manuscript writing and editing.

5.1 Abstract

Warmer temperatures in the Arctic have accelerated permafrost thaw both in depth and duration, threatening to release large portions of soil organic carbon (SOC) in the form of greenhouse gases (GHGs) like carbon dioxide (CO_2) and methane (CH_4). The amount of C released, and the proportion released as either CO_2 or CH_4, depends on many factors however, including temperature, hydrology, microbial community structure and function, and LMW DOM composition and availability. While the effects of temperature and hydrology have been studied extensively, the complex interactions between LMW DOM chemistry and soil microbial communities, and their effect on GHG production in response to thaw remains poorly understood. To take the first steps at addressing this knowledge gap, here we applied our untargeted LC/MS approach, that was developed in Chapter 3 and applied across multiple sampling sites in Chapter 4, to characterize LMW DOM along a natural permafrost thaw progression. Instead of the destructive harvest and aqueous extraction approach used previously, here we employed mini-rhizon samplers (see Chapter 2) to passively collect soil pore water *in situ* without disturbing the native soil structure. Using multivariate statistical analyses, features that were consistently and reliably detected, and were significantly differentially-abundant between sites, were annotated using high-mass accuracy MS measurements. Using our untargeted LC/MS approach, we provide a

detailed molecular profile of the shifts in LMW DOM availability in response to thaw-induced subsidence yielding mechanistic insight into how Arctic terrestrial systems may respond to continually warmer climatic conditions.

5.2   Introduction

It has been estimated that Arctic soils contain 1,400-1,850 petagrams (Pg) of carbon associated with soil organic matter (SOM), representing at least twice that found in the atmosphere (~800 Pg).[9, 15] Because the Arctic has historically acted as a net sink for atmospheric carbon, mobilizing and releasing even a fraction of SOC would represent a significant feedback to global climate change. In addition to predicting the amount of C that will be released, the proportion released as CO_2 or CH_4 is an important parameter in model predictions, primarily due to CH_4 having a 28-36 times higher global warming potential than that of CO_2 over a 100 y timescale.[218, 219] Whether SOC becomes CO_2 or CH_4 depends on many different factors including the hydrological conditions (*i.e.* anaerobic vs aerobic), the microbial community present (*i.e.* methanogens), and the type and availability of substrates (*i.e.* LMW DOM), all of which may be influenced by thaw-induced degradation. For example, permafrost is generally described as having two layers: an active layer that thaws seasonally and the permafrost layer below that, which remains frozen throughout the year. Over time, as active layer thickness increases (deepens) and previously-frozen permafrost soil and ice thaw, this can lead to inundated conditions and higher GHG emissions, especially methane.[220, 221] To explore these relationships and elucidate a chemical signature of C vulnerability in Arctic soils, here, we evaluate how LMW DOM composition varied along a natural thaw gradient.

5.3   Experimental approach

5.3.1   *Study site and sample description*

This study was carried out at the NGEE-Arctic field site established on the Seward Peninsula (64.89 °N, 163.67 °W), 57 miles east of Nome, AK on Council Road (Figure 45). The landscape is underlain by discontinuous permafrost and is characterized by heterogeneous ice distribution, well-defined watersheds, and large areas of thaw-induced subsidence.[222] This sub-Arctic site was chosen based on an analysis that indicated it to be a proxy for future ecological and climatic conditions on the North Slope of Alaska.[223] Soils consist of a wet (often saturated), thick (~20 cm), organic-rich surface horizon overlying mineral soil. Vegetation is fairly diverse, consisting of a mixture of mosses and lichens, grasses, woody shrubs, and even some trees, however this varied between sites along the thaw gradient (Figure 46). The annual mean temperature and precipitation for this region in May is 2.6 °C and 2.2 cm, respectively.[224]

Triplicate soil pore water samples were collected from three replicate natural thaw gradients (n = 9 per thaw condition) with three levels of degradation (27 total samples)—indicated here as "dry", "transitional", or "wet" (Figure 45). Samples were collected in May 2017. While all three sites were degraded, the "dry" area was the least degraded, and the "transitional" and "wet" areas were progressively more degraded due to permafrost thaw. While the "dry" and "transition" areas had a similar mixture of plant species, the "wet" site was dominated by grasses and moss. Mini-rhizon samplers were installed on Day 1 of field work, allowed to "equilibrate" for 24 hours, and collections were made on
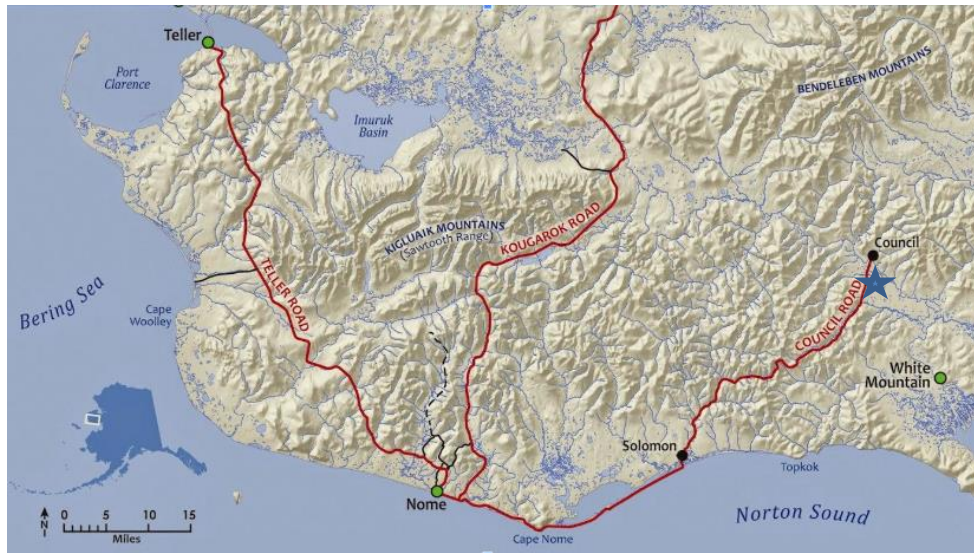
Figure 45: Map of Seward Peninsula showing Council road going East with location of the field site indicated by the blue star
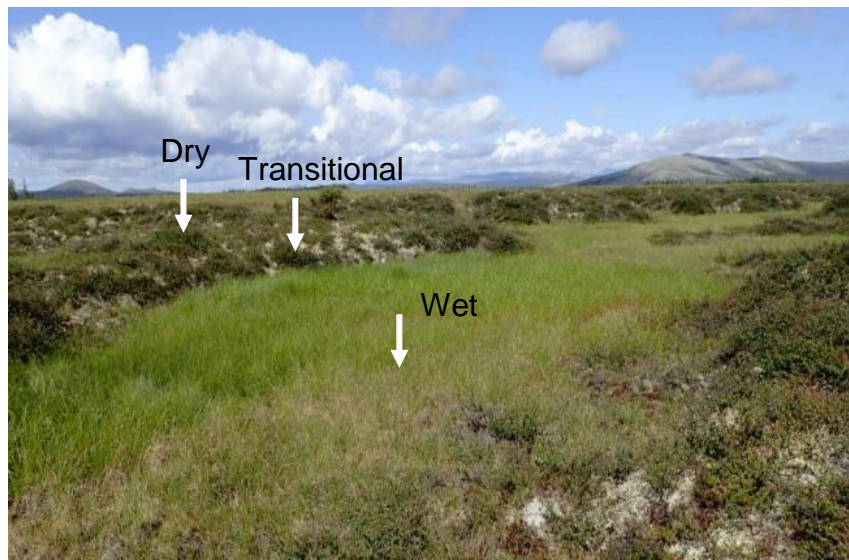
Figure 46: Photo of one of the natural thaw gradients located at the field site on Seward Peninsula near Council, AK; arrows indicate the three varying levels of degrdation (thaw) used in the study

Day 2. The vacutainers were immediately put on ice and kept frozen for shipment and storage (- 80 °C) until LC/MS analysis.

5.3.2   *Sample preparation and instrumentation*

Prior to LC/MS analysis, soil pore water samples were thawed and centrifuged to reduce probability of particulates that had passed through the rhizon being transferred to the autosampler vial and injected onto the column. One aliquot of each sample and control was evaporated down to ~ 5 µL and brought back to volume in 95 % ACN for HILIC analyses. Each sample was loaded into the autosampler and maintained at 4 °C. Nano-flow rates and 20 nL injections were achieved using a split-flow setup prior to the injection loop. Just as before, the pooled QCs were run every 6 injections and samples were randomized to reduce instrument-derived variation. Technical blanks representing the column re-equilibration conditions were also run regularly to monitor background ions and carry-over between samples.

Separations were performed using the same HILIC and RP-LC phases, setup, and optimized mobile phase conditions described in Chapters 2 and 3, resulting in four separate LC/MS analyses per sample (HILIC +/- and RP +/-, n = 108). Measurements of samples and controls were carried out using a Dionex UltiMate 3000 HPLC pump and autosampler coupled to a Q-Exactive Plus mass spectrometer. Here, the ESI source capillary temperature was optimized to 250 °C and the voltages for HILIC and RP were optimized to 3.4 and 1.5 kV, and 2.9 and 3.2 kV, for positive- or negative-ion mode, respectively. Full precursor (MS^1) scans were acquired in centroid mode at a resolving power of 70,000 over a mass range of 50 – 750 *m/z*. Fragmentation data were collected using collision-

induced dissociation (CID) with He$_{(g)}$ and was performed in data-dependent mode on the top 5 ions for each full scan at 35,000 resolving power, a 2 *m/z* isolation width, and 30 % normalized collision energy. Monoisotopic precursor ions that were selected for fragmentation were placed on a dynamic exclusion list for one minute and a charge state rejection of doubly-charged precursors was also enforced to improve detection and isolation of low abundant or coeluting, singly-charged small molecules. Two microscans were averaged for every full MS^1 and MS^2 spectrum to help reduce spectral complexity. Accurate *m/z* values were determined to four decimal places.

5.4 Results and discussion

5.4.1 C/N ratios decrease along natural thaw gradient

Along with LC/MS measurements, soil moisture data was collected in triplicate in the field just prior to sample collection using a 5TE soil moisture probe (Decagon Devices) and TOC and TN data were collected in triplicate on a Shimadzu TOC-L CSH/CSN analyzer back at ORNL, as described in Chapter 2. While the sites were indicated here as "dry," "transition," and "wet," it is important to note that the "dry" sites, on average, had a higher water content than the transitional areas (Figure 47). This may have been due to the close proximity of the transitional area to the wet area which was at a lower topography and caused significant pooling. The average C/N ratios of the soil pore water were highest at the least-degraded ("dry") sites (118 ppm +/- 97), intermediate in the transition zone (90 ppm +/- 56), and lowest in the "wet" areas (39 ppm +/- 18) where there was the highest amount of thaw-induced subsidence (Figure 47). This may have been due to the observed
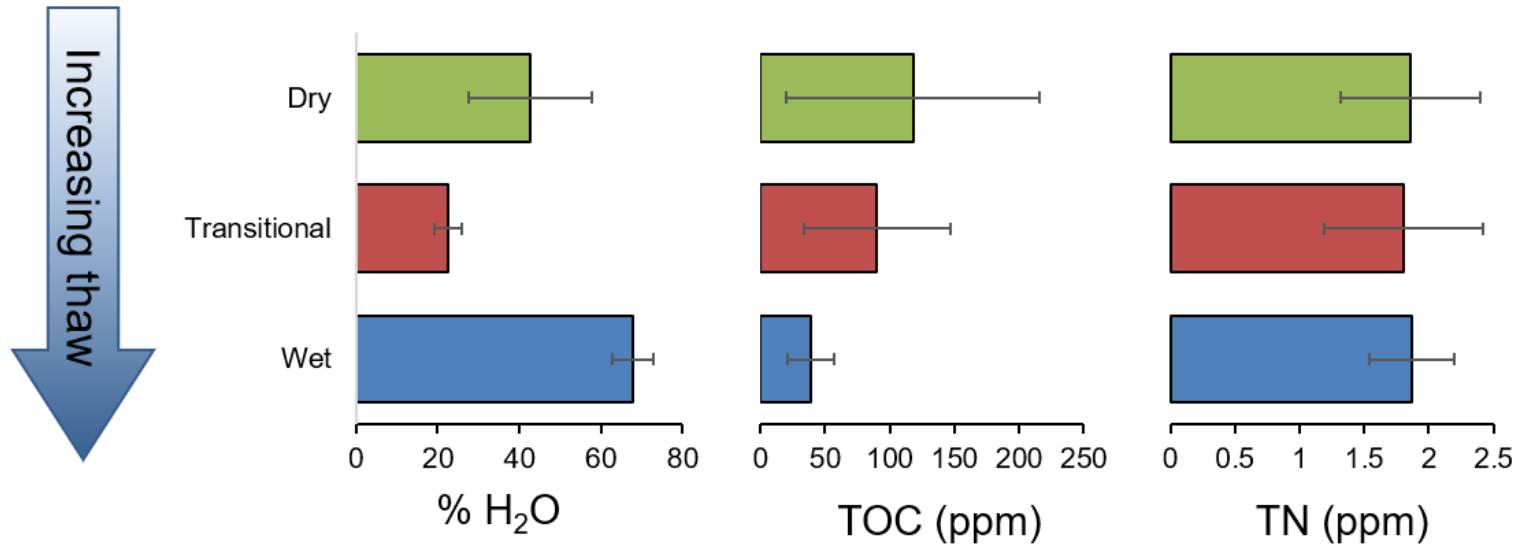
Figure 47: Soil moisture, total organic carbon, and total nitrogen measurements along the natural thaw gradient (top to bottom)

shifts in the plant community, consistent with the amount and/or quality of plant litter inputs varying along the thaw progression.

5.4.2 LMW DOM availability varies in response to thaw

Raw LC/MS files were processed with MZmine (v2.34) using the same steps described in Chapter 4, separated by LC/MS condition to avoid confounding effects between electrospray conditions. The results of the data filtering process from aligned peaks to annotated high-quality features is summarized in Table 13. Dissimilar from the first two studies where water extractions were used, more features were detected by RP (+) and RP (-) than the HILIC columns in these *in situ* pore water collections. However, of the 20,045 total peaks aligned, 10,395 (51.9 %) were detected in the blanks and negative controls (LC/MS-grade water through a rhizon) above the noise or abundance level thresholds and 68.7 % of these artifacts were from the RP conditions. After filtering these out, 9,313 high-quality features (HQFs) remained (Table 14).

Across the thaw gradient, more features were observed at the least-degraded sites than at the "wet," highly-degraded sites (Table 13) suggesting an accumulation of organic matter, consistent with the higher TOC contents observed at those sites (Figure 47). Using the unique IDs from each LC/MS condition, we assessed the overlap in HQFs between each of the sites along the natural thaw gradient (Figure 48). In general, there was a high degree of overlap between sites (64.4 %), with a higher amount of overlap between the "dry" and "transitional" sites (44.6 %) than either between the "dry" and the "wet" (17.5 %) or the "transition" and the "wet" sites (9.1 %). There were also unique features observed

Table 13: Total number of features detected across the three replicate gradients and carried through the data filtering thresholds for each LC/MS condition

	HILIC POS	HILIC NEG	RP POS	RP NEG
Peaks	2263	3801	9328	4653
Features	1027	1744	3750	3084
HQFs	1020	1728	3748	2817
Abundant HQFs	556	152	1691	1736
Differentially-abundant HQFs	257	91	394	882
Annotated	193	58	236	578

Table 14: Total number of HQFs detected at each site along natural thaw gradient by LC/MS condition

	Dry	Trans	Wet
HILIC POS	900	856	761
HILIC NEG	818	1048	464
RP POS	2892	3383	2489
RP NEG	2468	2565	2358
Sum	**7115**	**7852**	**6035**
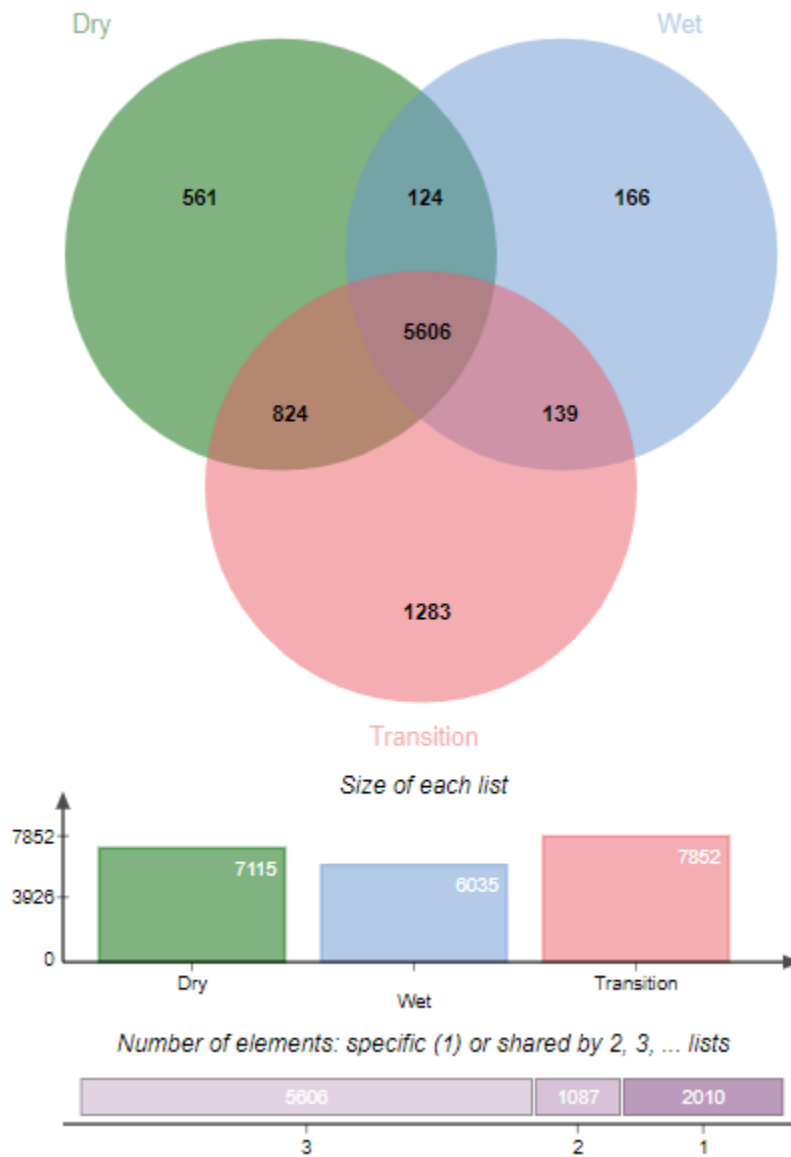
Figure 48: Overlap of HQFs, detected by all four LC/MS conditions, between sites along the natural thaw gradient

Venn diagram created with the jvenn tool.[225]

at each site (2 – 17 %), with the highest number of unique features observed at the transition zones, possibly due to the higher level of physical disturbance leading to more diverse soil environments (*i.e.* aerobic and anaerobic) and more diverse plant and microbial communities.

To investigate variations in the molecular composition of LMW DOM along the natural thaw gradient, features that were significantly differentially-abundant (ANOVA p-value < 0.05) between sites (dry vs transitional vs wet) were annotated by assigning molecular formulas taking into account C, H, O, N, S, and P and classified based on their H/C and O/C ratios, and DBE value (see Chapters 2 and 4). The total number of formulas assigned were 1,065 (65.6 %) out of the total 1,624 differentially-abundant features (Table 13). To visualize the differences in LMW DOM, we first plotted the compounds that were unique to each site based on the H/C and O/C ratios (Figure 49a) and N/C ratio (Figure 50) of the formulas assigned. Based on these data, in general, the variation in LMW DOM composition along the natural thaw gradient was minimal, possibly due to the early-season sample collection. As discussed in Chapter 1, spring can be a dynamic time in Arctic systems. These results indicate that early in the thaw season, there is an *increased* availability of LMW DOM across all three sites. This is consistent with previous studies that have shown there is a flush of organic nutrients upon snowmelt which commonly leads to the microbial community "waking up" before the plant community. This phenomenon is also supported by the ~ 70 % of formulas assigned here that contained N. Increased N availability relieves microbial N limitation, creating a smaller C/N imbalance, and altering microbial resource requirements, which has been shown to decrease microbial processing
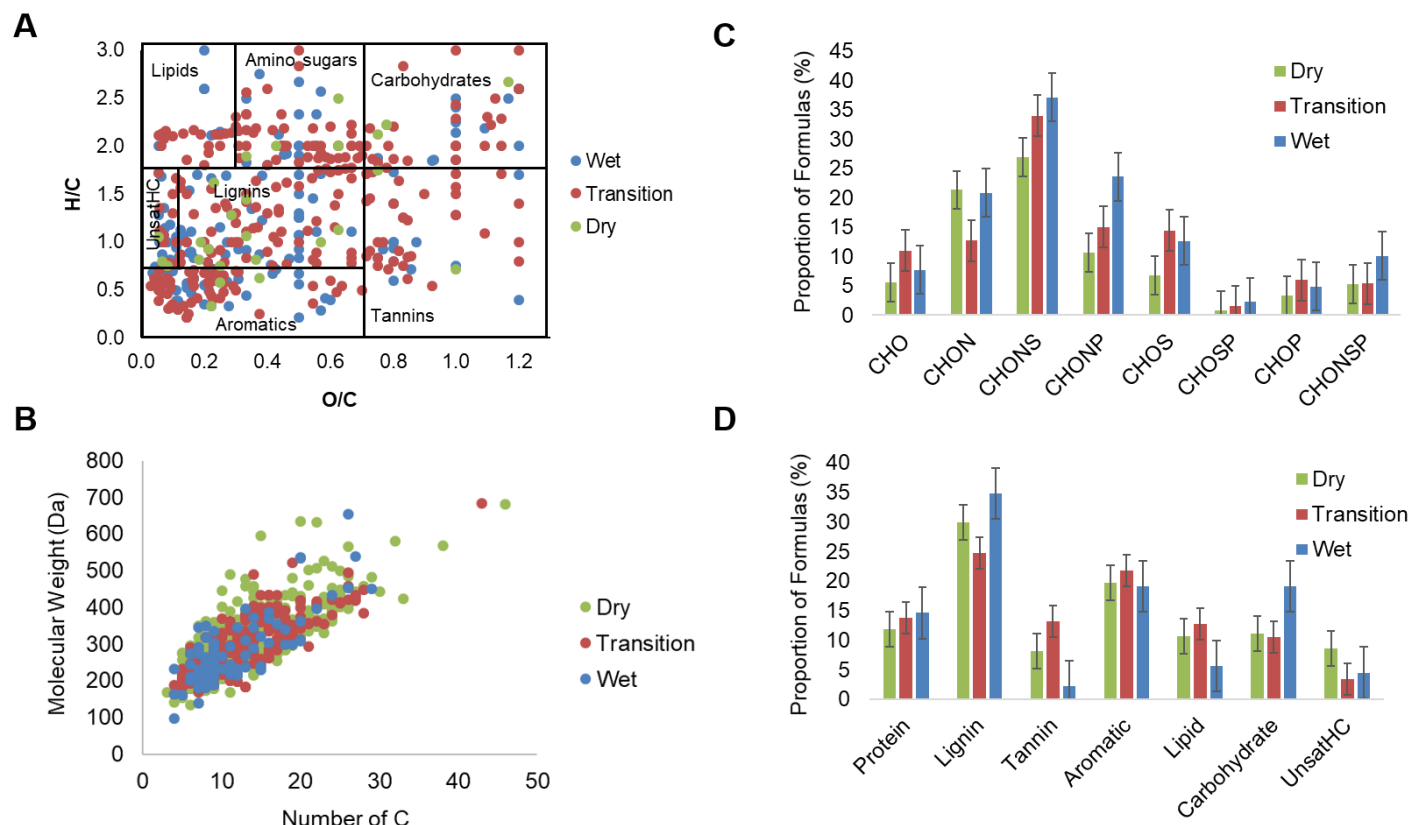
Figure 49: Molecular analysis of differentially-abundant (FC > 2, p-value < 0.001) LMW DOM features uniquely observed at each site along the natural thaw gradient. Relative abundance and distribution of different types of formulas assigned based on their (A) H/C and O/C ratios in a van Krevelen plot, (B) molecular weight, (C) elemental composition, and (D) compound class assigned.
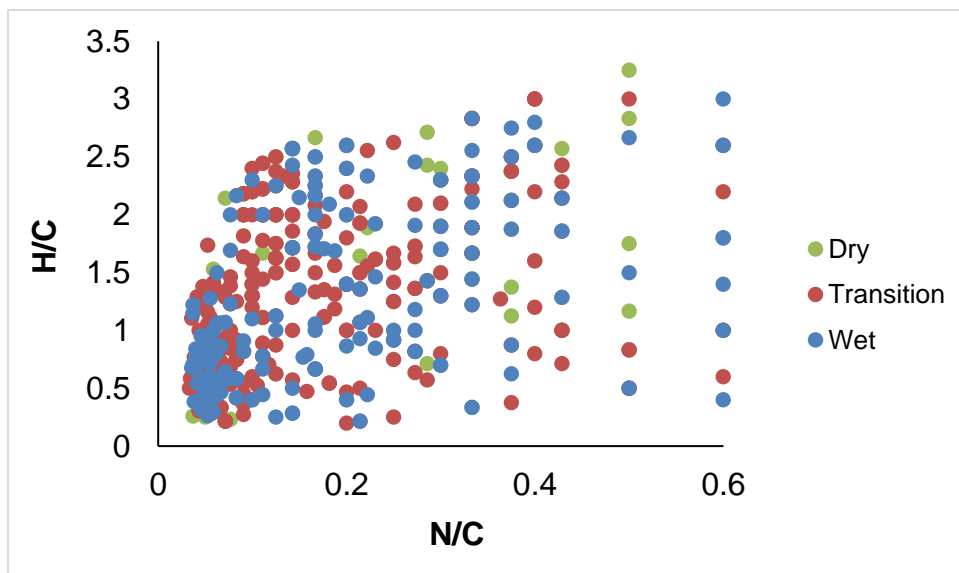
Figure 50: van Krevelen diagram of LMW DOM features that were differentially-abundant between sites along natural thaw gradient, using N/C ratio instead of O/C ratio

of organic matter as well as respiration rates.[34, 38] This is supported by the relatively low average C/N ratios at all three sites; 7.54 at "dry", 11.9 at "transition", and 9.45 at the "wet" sites.

Despite these similarities, there were some subtle variations detected reproducibly and robustly that indicate increased organic matter lability at the wet, more degraded sites. Since the degree of decomposition (*i.e.* relative abundance of structurally-complex molecules) is often inversely related to C/N ratios, the C/N ratios reported above support that the "transition" areas and the "wet" areas contained a LMW DOM pool that was slightly more vulnerable than that at the least degraded areas.[33] In addition, there were more formulas with lower molecular weights and higher N, P, and sulfur (S) content at the "wet" sites, indicating more advanced decay and a higher vulnerability to processing (Figure 49b-

c, Table 15). Furthermore, this was confirmed with the formula assignments consistent with lignin-type compounds being slightly higher at the "dry" sites, as well as tannin-like and aromatic compounds at the "transition" sites (Figure 49d, Table 16), both associated with decreased organic matter mobility and lower biodegradability due to a greater amount of energy required for microbial decay, although this is also temperature-dependent.[226, 227] Finally, enhanced LMW DOM vulnerability was also reflected in the degree of oxidation and unsaturation at each of the sites. Using the double-bond equivalents, we observed a decrease in the unsaturated aliphatic ring content and an increase in the level of oxidation (O/C and DBE/O ratio) with level of thaw (Figure 49d, Table 17).

Of the N-containing formulas, about half of those also contained S or P. Organic S and P in soil are largely immobile, as they are used by plants and microbes primarily for synthesizing amino acids and extracellular enzymes,[29, 228] and like organic N, are also limiting in Arctic environments. Only 1 – 3 % of microbial biomass is composed of organic S, but it is also the most readily available form of S due to the rapid turnover of microbial communities.[229] An accumulation of organic S, as indicated by these results, is consistent with an early-season flush of microbial cells that had turned over during the winter months. Likewise, while only a subtle difference, P increased with the level of thaw, indicated by a higher proportion of CHOP, CHONP, CHOSP, and CHONSP formulas at the "wet" sites (Table 15, Figure 49c). This is consistent with previous studies that have shown a relationship between P availability and hydric stress.[230] Along the natural thaw gradient here, an increase in water availability may have led to an increase in phosphatase activity, ultimately leading to more available organic P for microbial decomposition.

148

Table 15: Proportion (%) of formulas with distinct elemental compositions (CHO only, CHON only, etc.)

	CHO	CHON	CHONS	CHONP	CHOS	CHOSP	CHOP	CHONSP
Dry	5.6	21.3	27.0	10.6	6.7	0.8	3.4	5.3
Transition	11.0	12.7	34.0	15.1	14.4	1.5	6.0	5.4
Wet	7.8	20.8	37.1	23.6	12.7	2.2	4.9	10.1

Table 16: Proportion (%) of assigned formulas belonging to each compound class detected distinctly at each site along the thaw gradient

	No. of Formulas	Protein/ Amino Sugar (%)	Lignin (%)	Tannin (%)	Aromatic (%)	Lipid (%)	Carbohydrate (%)	Unsat HC (%)
Dry	89	14.6	34.8	2.2	19.1	5.6	11.1	8.6
Transition	465	13.8	24.7	13.1	21.7	12.7	10.5	3.4
Wet	245	11.9	29.9	8.2	19.7	10.7	19.1	4.5

Table 17: Average oxidation states and degree of unsaturation at each site along the thaw gradient

	O/C	H/C	DBE	DBE/O
Dry	0.421	1.444	7.231	2.424
Transition	0.459	1.314	7.111	2.601
Wet	0.463	1.300	6.174	2.803

5.5   Conclusions

Taken together, these results indicate a high degree of overlap in the LMW DOM chemistry between sites, possibly due to an early-season flush of microbial biomass and organic nutrients upon snowmelt that accumulated over the winter months. However, there were also measurable differences in the LMW DOM availability across the thaw progression, with a slight increase in organic matter vulnerability at the more degraded sites supporting that LMW DOM may be an important source of GHG emissions from thawing permafrost soils. While most measurements of this kind have been completed in the laboratory with destructive harvests and incubation analyses, here we provide a detailed molecular profile of LMW DOM availability collected *in situ* using passive pore-water samplers. We propose that this information-rich chemical fingerprint could be correlated with co-located measurements of GHG emissions, plant community composition, and/or microbial community structure and activity to elucidate a profile of compounds that could serve as specific markers of C vulnerability, yielding insight into the underlying, complex mechanisms that control C sequestration or release in Arctic systems.

CHAPTER 6: ARCTIC CLIMATE POLICY ASSESSMENT

6.1 Abstract

A unique and important aspect of the Energy Science and Engineering doctoral program is the incorporation of an interdisciplinary component. Thus, the following chapter critically evaluates three U.S. policy alternatives for addressing the impacts of climate change in the Arctic, including an analysis of how -omics technologies can inform Arctic science and policy. I employ a logical assessment approach that first identifies a policy challenge, proposes various alternatives to address that challenge, lists objectives that the proposed policy alternatives should meet, and then compares the alternatives based on their technical, political, and economic feasibility. Where appropriate and when data were available, quantitative methods of comparison were employed to assist with maintaining an objective analysis. While this is not an exhaustive comparison of all policies that have been proposed for addressing the impacts of climate change in the Arctic or their outcomes, the narrowed list here were chosen after a review of current and relevant literature; they represent not only a diverse range of approaches but are also some of the most commonly-discussed among experts in the field.

6.2 Introduction

The United States purchased the territory now known as Alaska from the Russian Empire nearly 150 years ago, officially making it one of the eight Arctic nations. For much of its history however, our corner of the Arctic was often left out of U.S. policy discussions and was considered too remote for scientific exploration. In recent decades however, with warming temperatures causing rapid environmental change, and an ever-increasing human population creating a growing need for more energy and natural resources, our attention

152

has turned north both in terms of science and policy. Rising temperatures have led to sea-ice recession, declining snow cover on land, and increased areas of frozen ground (permafrost) beginning to thaw. This has the potential to increase access to natural resources such as oil and gas reserves both on- and offshore and open new shipping routes, increasing trade and commercial activity in the region.

However, there are many challenges associated with these changes as well. Increased human activity in the region creates a need for increased security and development of protocols for spill-response and search-and-rescue missions for example. In addition, coastal erosion and destabilization of the permafrost has already had devastating effects on local infrastructure (Figure 51). Foundational settling due to permafrost thaw can destroy buildings, roads, pipelines, railways, and power lines resulting in substantial maintenance and repair costs.[231] Furthermore, Alaskan ecosystems are already experiencing changes in plant species composition, animal migration patterns, and increased intensity and frequency of forest fires, all of which have significant social, cultural, health, and economic impacts on local human populations (*i.e.* food security, ecosystem biodiversity).[232, 233] Finally, as has been discussed in previous chapters, thawing permafrost also leads to carbon-rich organic matter suddenly becoming available to microbial decomposers, where it can then be released from the soil in the form of greenhouse gases (*e.g.* carbon dioxide, methane) creating an irreversible feedback to the global climate system.

Figure 51: Images showing infrastructure damage due to thawing permafrost and eroding coastlines in the Alaskan Arctic

Clockwise, from top left – 1) Exit Glacier Rd during high water event in 2009 2) flood waters rushing over Exit Glacier Rd in 2010, 3) maintenance at hillside slump at Mile 20.5 of Denali Park Rd in 2005, and 4) eroding shorelines and thawing permafrost leading to infrastructure damage at Bering Landbridge National Preserve. Source: Public Domain – National Park Service (NPS)[234, 235]

6.2.1 Policy challenge statement

The Alaskan Arctic has become a region of national significance, as rapid climate-driven change has led to cascading effects on the environment, human health, infrastructure, energy development, and national security. As an Arctic nation, the United States has a responsibility to act on these impacts.

6.2.2 Existing policy framework

To appropriately evaluate a range of policy alternatives that could feasibly help address this challenge, a review of existing and relevant legislation, the objectives those pieces of legislation were founded upon, and U.S. administrative bodies that direct those policy initiatives was first conducted.

Although the U.S. has owned the Alaskan territory since 1867, the Nixon administration was the first to mention a set of policy priorities for the region in 1971 that focused on three key areas:[236]

- Minimizing risks to the environment

- Promoting international cooperation, and

- Protecting security interests in the region

Since then, although more detailed initiatives have been put forth, U.S. priorities in the Arctic remain strikingly similar, and have a strong foundation of including scientific research. In 1980, under President Carter, the Alaska National Interest Lands Conservation Act (ANILCA) was passed providing protection to over 100 million acres of land in national parks, wildlife refuges, monuments, wild and scenic rivers, recreational and

conservation areas, and forests to assist with scientific research, increase tourism, and reduce the impacts of commercial activities.[237] In 1984, under President Reagan, the Arctic Research Commission was created as a part of the Arctic Research and Policy Act (ARPA) to establish national policy and research priorities in the region.[238] Then, in 1991, the Arctic Environmental Protection Strategy agreement with the other seven Arctic nations (Canada, Denmark, Finland, Iceland, Norway, Russia, and Sweden) was initiated eventually leading to the creation of the Arctic Council in 1996, formally including the U.S. in international Arctic science and policy decision-making. In 2009, the U.S. government under the Bush administration released the National Security Presidential Directive (NSPD-66) setting a more expanded list of priorities for the Arctic, including:[239]

- Meeting national security and homeland security needs relevant to the Arctic region
- Protecting the Arctic environment and conserving its biological resources
- Ensuring that natural resource management and economic development in the region are environmentally sustainable
- Strengthening institutions for cooperation among the eight Arctic nations
- Involving the Arctic's indigenous communities in decisions that affect them, and
- Enhancing scientific monitoring and research into local, regional, and global environmental issues

This was an important acknowledgement of several areas that needed further research and formalized policy options, including from social, environmental, and security perspectives. Then, in 2013 under the Obama administration, a somewhat more detailed strategy was developed outlining several specific objectives, which included:[240]

- Evolving the Arctic infrastructure and strategic capabilities

- Enhancing Arctic domain awareness

- Preserving Arctic freedom of the seas

- Providing for future U.S. energy security

- Protecting the Arctic environment and conserving natural resources

- Using integrated Arctic management to balance economic development, environmental protection, and cultural values

- Increasing understanding of the Arctic through scientific research and traditional knowledge

- Charting the Arctic region

- Pursuing arrangements that promote shared Arctic state prosperity, protection of the Arctic environment, and enhanced security

- Working through the Arctic Council to advance U.S. interests in the Arctic, and

- Acceding to the United Nations Law of the Sea Convention (UNCLOS)

Over the years, multiple working groups, task forces, and various federal offices and agencies have been involved in implementing U.S. science and policy strategies and objectives including most prominently the U.S. Arctic Policy Group chaired by the Department of State, the Interagency Climate Change Adaptation Task Force initiated by the Office of Science and Technology Policy in the White House, and the U.S. Navy's Task Force on Climate Change under the Department of Defense. In addition, the National Science Foundation, the Interagency Arctic Research Policy Committee, the Arctic Research Consortium of the United States, the Environmental Protection Agency, and the

Departments of Commerce, Interior, Homeland Security, Energy, Agriculture, and Transportation have all contributed to various initiatives in recent years that have helped promote the development of Arctic science and policy to address the impacts of climate change in the region. Beyond the federal government, there are also many other avenues by which Arctic science and policy agendas are developed, coordinated, and implemented including the Arctic Council, bodies within the United Nations, regional initiatives such as the International Polar Year, professional associations like the International Arctic Science Committee, and international conferences. Additional stakeholders who have a vested interest in Arctic science and policy are included below in the Political Feasibility section.

6.2.3   Objectives

From this review of existing legislation and recommendations made by various stakeholders, the following six objectives, in no particular order, were most commonly observed, and were used here to evaluate the proposed policy alternatives:

1) Promote scientific and political cooperation in the Arctic region

2) Protect the Arctic environment

3) Optimize access to energy resources

4) Maximize political feasibility

5) Minimize social and economic impact on Alaskan citizens

6) Minimize policy-implementation costs to the federal government

Each of the policy alternatives proposed here (below) will also be reviewed based on their technical, political, and economic feasibility. To help objectively assess each of

these areas, a quantitative comparison will be employed by assigning a numerical weight to each objective or a feasibility score based on a review of publicly-available documents and statements made by various stakeholders.

6.2.4 *Proposed policy alternatives*

Two frequently-proposed policy alternatives that address several of the objectives listed above were selected to be evaluated here first:

Alternative 1: Ratify the United Nations Convention Law of the Sea (UNCLOS)

The UNCLOS treaty is an international agreement that was formulated between 1973 and 1982 with the aim of defining the rights and responsibilities of nations for how to conduct business and science on the world's oceans, protect the marine environment, and manage marine natural resources.[241] While the U.S. remains party to the provisions in the Convention, and there has been broad and continued support through multiple administrations, it has not yet been formally ratified into law. Formally joining the treaty would not only encourage international cooperation but provide a formal legal framework for defining off-shore land and resource claims—a common cause of maritime disputes that arise in the Arctic region between Arctic and non-Arctic nations alike. The treaty would also enable these maritime disputes to go to tribunal for diplomatic resolution. Joining the convention would directly enhance political and military international cooperation and may also indirectly promote scientific cooperation to expand charting and/or conservation efforts.

Alternative 2: Increase the amount of federally-protected land under the Alaska National Interest Lands Conservation Act (ANILCA)

Currently protecting 104 million acres of Alaska's land, ANILCA passed into U.S. law in 1980, as outside interests increasingly sought to export Alaska's oil, fish, timber, and minerals for profit. With ice and permafrost thaw creating new opportunities for energy development and resource exploration (*e.g.* drilling and mining), it has been suggested that increasing the amount of federally-protected land may be necessary once more and would reduce the risk of further permafrost degradation and the subsequent impacts on local wildlife or indigenous communities for example. Although somewhat unpopular when first set into law, the conservation and economic benefits of ANILCA have generated strong support over time.

In addition to these two policy alternatives, because scientific research is a core principle of U.S. policy strategies for the Arctic, it has been proposed that establishing robust, sustainable, and participatory research networks would not only enable international cooperation but result in more sound and actionable policies based on evidence generated from scientific studies and analyses.

Alternative 3: Establish an integrated -omics research network for Arctic science

One such example of this, that would facilitate in providing detailed and holistic insight into Arctic ecosystem health and function, is the creation of a research network that focuses on generating, integrating, maintaining, and disseminating data, methods, and results from genomics, transcriptomics, proteomics, and metabolomics (*i.e.* integrated -omics) experiments (see Underlying Science section for more detail).

160

While several policy actions have been recommended previously in the public sphere, and common objectives exist across those recommendations, few have been evaluated side-by-side, or in a quantitative manner. Here, I evaluate these three contrasting policy alternatives in terms of their technical, political, and economic feasibility, and ultimately make a policy recommendation based on the results of this comparative analysis.

6.3   Underlying science

While not always employed in current legislative procedures, a critical component of any public policy decision-making process should be a review of the relevant scientific information available to ensure any anecdotal or historical arguments are supported by empirical evidence. Indeed, organizations like the American Association for the Advancement of Science and others have become more vocal about this in recent years.[242] Thus, to ensure a comprehensive analysis of the three policy alternatives examined here, the underlying scientific data that may inform the decision-making process was first reviewed.

Alternative 1: Ratify UNCLOS

A core component of UNCLOS is that it establishes a legal framework for the definition of territories currently unclaimed in the Arctic seas. Each Arctic nation has an exclusive economic zone (EEZ) which extends 200 nautical miles from shore. However, the U.S. continental shelf—the shallow continuation of our land mass deep under water— extends beyond the EEZ where it transitions to the deep ocean floor.[243] UNCLOS confirms that coastal states have sovereign rights over the sea bed and subsoil beyond the EEZ but

that each nation must define their extended continental shelf and that these scientific data must be submitted to UNCLOS for validation.

Since 2007, U.S. agencies have been engaged in collecting and analyzing data to define the boundaries of the U.S. continental shelf off the coast of Alaska. The primary means by which scientists are helping fill this knowledge gap are with bathymetric surveys. Bathymetric surveys use multibeam sonar, a type of sound transmitting-and-receiving system that sends a pulse at a specific frequency toward the sea floor, and then determines the time it takes to receive the returning signal, which can then be translated to depth and used to create three-dimensional images of the sea floor.[244] With more waves bouncing back, more accurate and higher-resolution images can be created. Despite this, an additional challenge that exists in the Arctic is that much of the continental shelf is hidden below a thick layer of sea ice most of the year. With just two U.S. ice-breakers in commission, even though the Arctic Ocean is the smallest and shallowest of the five major oceans, just 2.5 % of it has been surveyed with modern methods and technology.[245]

Alternative 2: Increase the amount of federally-protected land under ANILCA

Currently, ANILCA protects 104 million acres which includes national parks and preserves, wildlife refuges, wilderness areas, wild and scenic lakes and rivers, and the Iditarod National Historic Trail comprising approximately 24.5 % of the total area of Alaska. A key distinction between federally-protected land in the "lower 48" states and that of Alaska is that provisions have already been made allowing for subsistence hunting and fishing, public use of cabins or shelters, and use of snowmobiles, motorboats, and airplanes in these areas, as well as sport hunting in the wilderness preserves.[246] It is also

important to note that, while viewed as short-sighted in the details early on, over time and with a few updates, ANILCA was incredibly forward-thinking in what it would ultimately accomplish with respect to conservation and supporting the local economy (see economic feasibility analysis). Because of this, ANILCA is generally seen as a positive initiative despite some resistance from a few lawmakers (see political feasibility analysis).

Alternative 3: Establish an integrated -omics research network for Arctic science

Central to any Arctic science or policy agenda is increasing our capacity to model the impacts of climate change across many different scales and identify trends and indicators for future change in climate, human health, and biodiversity. There have also been calls for a better fundamental understanding in the areas of shifting food web structures, enhanced competition between plant or animal species, increased predation, and shifting population dynamics like changes in size or structure in response to varying habitat conditions or pollution levels for example.[233] With the advent of highly sensitive, nontargeted analytical technologies, more and more fields, including energy and earth system sciences, have been turning to *omics* approaches for a more detailed, mechanistic understanding of how biological systems function.

The term *omics* refers to various fields of study that seek to describe a biological system by characterizing and quantifying biological molecules that yield insight into the collective structure and/or function of that system. At the top of the "omics cascade," and one of the oldest omics fields, is that of *genomics*, where an organism's DNA is studied in order to map its genome ultimately identifying "who's there." *Transcriptomics* and *proteomics* are next in line, as they describe the RNA and proteins, respectively, that are

produced by the genome, giving more insight into "what that organism has the capacity to do." While these technologies have become highly sensitive and more routine in being able to describe the structure of an organism or community, function cannot be inferred through genomics, transcriptomics, or proteomics approaches alone. *Metabolomics* however, is the comprehensive analysis of all metabolites within a biological system and thus represents a snapshot in time of "what the organism has done," yielding insight into function and/or phenotype. There are also other omics techniques that focus on a specific class of compounds (*i.e.* lipidomics), the transformation of molecules through multiple processes (fluxomics), or larger systems that look at multiple organisms and how they interact with each other or their environment (*i.e.* interactomics, exposomics).

While each individual field has been around for quite some time and each has seen substantial growth over the years, the combination of omics tools—*integrated omics*—has only recently been receiving increased attention (Figure 50).[247] This is primarily due to technical advancements in collection of the omics data and its newly realized potential to yield insight into both structural and functional attributes of a system simultaneously.[248-251] In addition, the bioinformatic tools (i.e. statistical machine learning, high-performance computing) capable of processing and analyzing these large datasets have also seen substantial growth alongside the data collection technologies.[252-255]

There have even been a handful of studies in the Arctic that have used integrated omics approaches to research various biological systems and their impact on different processes including: carbon cycling by microbial communities in permafrost[256, 257] or snow,[258] the efficiency of native soil microbial communities at degrading diesel fuel,[259] or
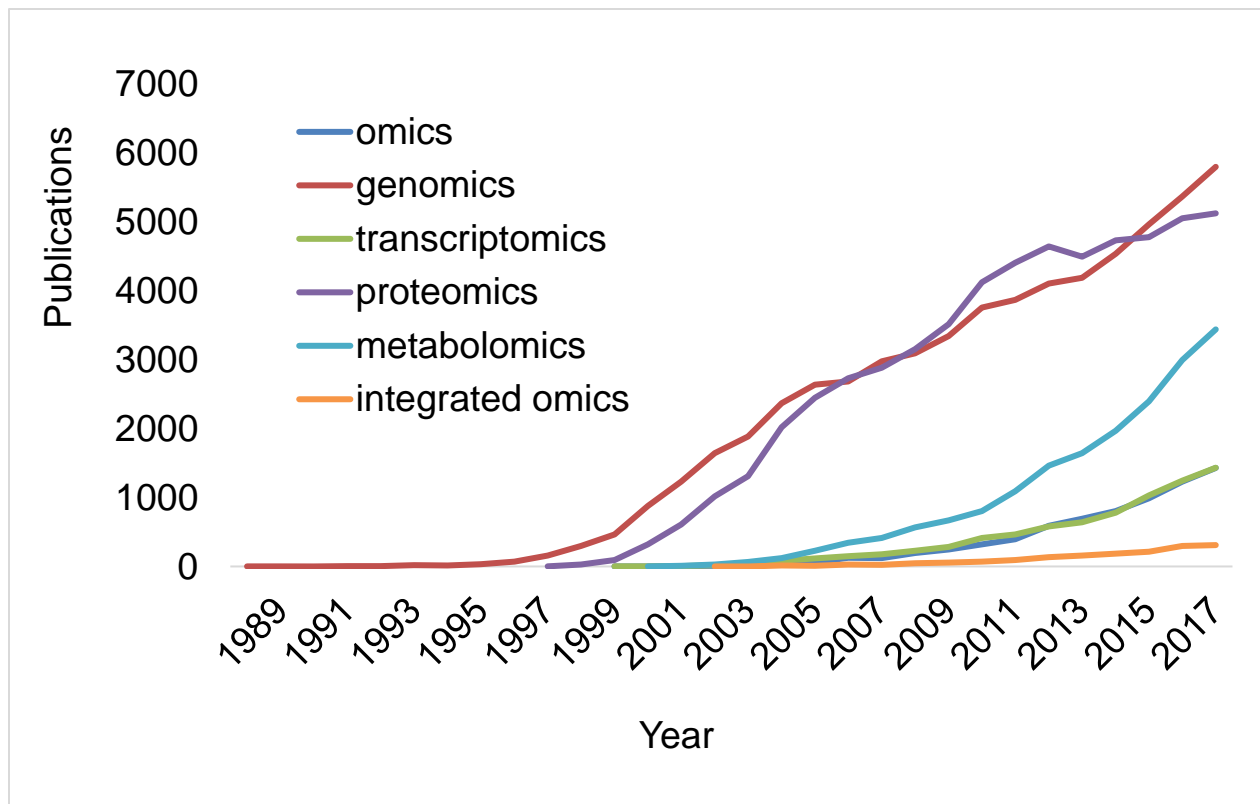
Figure 52: Keyword analysis of published, peer-reviewed journal articles on Web of Science that used omics techniques individually or integrated omics

the functioning of microbial communities in wastewater treatment systems to determine how their structure and function impact efficiency in a polar setting.[260] Whether to study climate change, bioremediation, or human health, integrated omics technologies have been established as having a great potential to help capture the multifaceted responses of Arctic ecosystems to warming temperatures and increased economic activity.

6.4   Feasibility analysis

In addition to evaluating the underlying science that would inform policy development and implementation, a key part of the decision-making process is to assess the technical, political, and economic feasibility of the policy alternatives. Technical feasibility refers to whether the appropriate technologies necessary to implement the policy exist, are readily (often, affordably) available, and if the measurement being made reaches a desired level of reliability. When no technology is necessary, a technical feasibility analysis will often include an evaluation of the effectiveness of the proposed policy using previous similar policies or other countries' policies as case studies. Political feasibility refers to the extent that a proposed alternative will be acceptable to various stakeholders. Finally, economic feasibility describes an analysis of the impacts a proposed alternative would have on the economy, whether that be at the federal or local level, in the public or private sector, or even at the level of the individual household.

6.4.1   Technical feasibility
Alternative 1: Ratify UNCLOS

While ratifying the UNCLOS treaty does not directly require any technology, much of the reason for why it has not yet been signed is due to claims that obligatory international

cooperation would infringe on U.S. sovereignty in the region and limit possible economic growth, claims that stem from a lack of detailed, technical data related to charting territories and energy reserve estimates. For example, by accurately charting the area of the Arctic Ocean that is U.S. territory, we would enter into this agreement with more knowledge and readiness to appropriately advocate on behalf of U.S. interests in maritime disputes. While the technology to map the extended continental shelf is readily available, there is some uncertainty associated with bathymetric measurements and some challenges that still exist. For one, while warming temperatures are indeed causing sea ice recession, for most of the year there still exists a thick layer of ice making it difficult or impossible for bathymetric surveys to be conducted without an ice breaker, and the U.S. has just two of those currently. Second, regarding the accuracy of these measurements, the greatest areas of uncertainty occur nearest to shore, on steep slopes, and near the range limit of the sonar (i.e. deep canyons), when the waves pass through an object instead of bouncing off it. However, the U.S. Army Corps of Engineers along with the National Ocean Service has created a set of minimum accuracy standards that quantify the error in these measurements and identify bounds for reliability to support charting and planning activities.[261]

Another significant scientific unknown that many stakeholders have called for more information on is accurate estimates of the amount and value of the oil, gas, and minerals that lie beneath the extended continental shelf. This is completed by geologists that identify areas with the right conditions for drilling and/or extraction (*i.e.* source rock, soil types, level of entrapment). Satellite imagery and gravity meters are both used to examine sub-ocean terrain and detect small variations in Earth's gravitational or magnetic fields that

could indicate flowing reserves of oil for example. In addition, infrared and thermal imaging has been used to detect hotspots of hydrocarbon release. Finally, seismology is the most commonly employed technique, where various types of shock (sound) waves are created—using compressed-air guns, thumper trucks, or explosives—passed through hidden rock layers, and then reflected back to the surface and detected by hydrophones.[262] While the technologies to estimate energy reserves in the Arctic ocean exist, the same environmental (*i.e.* sea ice, freezing temperatures) and technical (*i.e.* measurement feasibility and uncertainty) challenges described above have to be considered here as well. Indeed, in a recent review, it was found that estimating reserves is "highly uncertain" and there exists a "large degree of variability"[263] suggesting technical advancements are still needed before this policy alternative may be politically feasible.

Alternative 2: Increase the amount of federally-protected land under ANILCA

While no technology is needed to increase the amount of federally-protected land in Alaska, one of the most prominent reasons that has been argued to do so is for conservation purposes. As such, a brief review of the literature concerning the effectiveness of protecting land for conservation is reported here.

Most evaluations of whether protecting land was effective for conservation efforts (*i.e.* reducing deforestation, maintaining or improving biodiversity) rely on comparing protected areas to unprotected areas. However, it has been shown that where protected areas are placed is biased in nature, and surprisingly, biased toward areas that are unlikely to face land conservation pressures.[264, 265] In a recent analysis that used unprotected public and private (managed) lands as the controls, significant differences were still found

168

between protected and unprotected lands with respect to deforestation (less deforestation at protected sites).[266] Additionally, in a study that sought to evaluate the effectiveness of protected areas for conserving biodiversity of species that are changing their geographic distributions due to climate change, there was still a positive effect observed in bird and butterfly populations on protected lands for preventing extinction and promoting colonization.[267] Conversely, newly protected areas that thereby promote increased levels of tourism have been shown to negatively impact certain mammal species-richness and correlate with a decline in overall populations.[268] With mixed responses such as these, an evaluation of the effectiveness of conservation efforts by protecting land specifically in the Alaskan Arctic is warranted.

Alternative 3: Establish an integrated -omics research network for Arctic science

While full genome sequencing has become a routine analysis, and state-of-the-art proteomics and metabolomics technologies have become more affordable, available, and technologically-sound (*i.e.* better coverage, fewer false positives), integrating omics is still a relatively new area of study with some challenges. The technological capability to generate these large datasets is well-established and the uncertainty in those measurements gets better each year, but the ability to extract biological information from the hundreds of thousands or even millions of data points, remains a challenge in the field. However, no single omics science by itself can obtain a comprehensive understanding of a biological system, and as such, the field of integrated omics will only continue to grow. By conducting a review of relevant literature,[248, 251, 254, 269-274] some of the main challenges surrounding integrating omics technologies have been summarized below:

- methodological variation; lack of reproducibility due to nonuniformly standardized sample preparation, data collection or entry across existing databases; and rapidly evolving analytical technologies frequently give rise to new types of data creating a need for new data analysis platforms as well;

- need for high-throughput, multiplexed approaches (parallel measurements);

- scaling-up omics measurements to yield meaningful insight into larger-scale research questions (*e.g.* climate science), transitioning from observations to management applications (*e.g.* bioremediation, synthetic biology);

- ensuring omics results have added value to existing paradigms of Arctic science; whether they only add incremental value to current policy decision-making protocols needs to be evaluated;

- informatics challenges: more mature mechanistic models, data storage limitations, and organization/combination of fragmented databases or datasets; and

- disseminating, managing, and interpreting omics data in a broader policy context.

To summarize the results of this technical feasibility analysis, I have listed each alternative below with an assigned score out of 10 for each category based on the discussion above (Table 18). A higher score corresponds to a higher feasibility for that proposed alternative.

Table 18: Summary of technical feasibility scores assigned to each policy alternative

	UNCLOS	ANILCA	OMICS
Technology exists	8	10	8
Readily available (affordable)	6	10	7
Reliable/Accurate	5	10	7
Effectiveness	8	6	8
Total	**27**	**36**	**30**

6.4.2 *Political feasibility*

In addition to technical feasibility, one fundamental criterion legislators use to evaluate the likelihood of success is political feasibility. With, frequently, multiple stakeholders involved, each with their own perspective and set of beliefs or motivations, determining the overall political efficacy of a proposed policy can quickly become a complex process. For example, some of the perspectives that may come into play in determining Arctic policy are groups that would be interested in our national energy strategy as well. These varying energy perspectives were described recently in detail elsewhere[275] and have been reproduced here (Table 19) in terms of what goals they may consider when determining which of the Arctic policy alternatives best suits their missions.

Identifying these varying perspectives helps generate a diverse and balanced list of stakeholders both for and against the proposed alternatives. Groups that were considered here include various government agencies, private industries, academic institutions, nongovernmental organizations (NGOs), and citizen's groups. The stakeholders identified for this analysis are summarized below in Table 20 and a brief discussion of public comments they have made about each of the alternatives follows.

Table 19: Political perspectives of a diverse range of energy policy stakeholders

Perspective	Primary Goals
America-firsters	Energy independence, large military presence, security
Bottom-liners	Secure, low-cost national energy portfolios
Entrepreneurs	American market-place ingenuity
Environmentalists	Reduce greenhouse gas emissions, conservation
Individuals	Maintain high quality of life
Politicians	Accommodate many interests
Technophiles	Advocates of "big engineering" to achieve energy independence and lower greenhouse gas emissions

Table 20: Summary of stakeholder groups used to contrast policy alternatives in political feasibility analysis

Stakeholder	Values
Alaska residents	Quality of life, financial gain, energy efficiency, low fuel costs, land access, preservation of Arctic
Republicans	National defense, low government spending, job creation, sovereignty
Democrats	National defense, sustainable energy and environmental policy, social equity
Labor unions	Maximize benefits, quality of life
Academics	Scientific research and understanding
DOE/EPA/NSF/DOI	Scientific discovery, human health and environment, energy security, Arctic health, civil infrastructure
AK Dept. of Natural Resources	Protect and enhance natural resources; social equity
Bureau of Land Management/Fish & Wildlife	Sustainable health, diversity, and productivity of public lands, tourism
Gas/petroleum/minerals industries	Minimize costs, maximize profits and economic penetration
U.S. Navy/Dept. of Defense	National defense, coastal and marine security, scientific research and discovery

Alternative 1: Ratify UNCLOS

When asked about whether she supports U.S. accession to UNCLOS, Alaska Senator Lisa Murkowski (R – AK) has been quoted in saying "it is crucial for the United States to be a party to this Treaty."[276] Just as recently as July of 2018, Senator Murkowski introduced a bipartisan bill with Senator Hirono (D – HA), with support from their constituents, urging approval of UNCLOS.[277] There have been a few Republican senators, like James Inhofe of Oklahoma for example, that view the treaty as "a threat to U.S. sovereignty,"[276] but most lawmakers have held similar views to that of Senator Murkowski, like Senator Hank Johnson (D – GA) for example, who through discussions with energy groups like the American Petroleum Institute, environmental groups like the Ocean Conservancy and the World Wildlife Fund, and labor unions such as the American Federation of Labor and Congress of Industrial Organizations and the Seafarers International Union of North America, have found broad support for the ratification of UNCLOS, as it would "extend American interests…beyond the 200-nautical mile Exclusive Economic Zone" so that "American businesses can develop and invest in maritime resources…knowing they are supported by the legal certainty and stability of treaty law."[278] Similarly, academics in the areas of law and economics alike have also found that the costs of *not* joining UNCLOS outweigh the potential benefits.[279, 280] Finally, the Department of Defense—the U.S. Navy in particular—has been a very vocal supporter of ratifying UNCLOS. In 2000, the retiring Chief of U.S. Naval Operations, Admiral Jay Johnson wrote to the Senate Foreign Relations Committee that he "consider[ed] UNCLOS [his] most significant piece of unfinished business,"[281] and just as recently as 2015, the

Department of Defense was quoted in saying about UNCLOS that "adherence to a rules-based system has been critical to furthering peace, stability, and prosperity."[282]

Alternative 2: Increase the amount of federally-protected land under ANILCA

Although it is frequently proposed that the U.S. increase the amount of land protected under ANILCA, it has also frequently been met with resistance from various groups including oil and gas industries and Republican lawmakers, like Senator Murkowski of Alaska who recently stated the "federal government…is trampling on [Alaska's] state sovereignty" with how ANILCA is being implemented.[283] Likewise, some Alaskan citizens have stated that expanding ANILCA would "force [them] to live in a permit society," without access to the timber, oil, gas, and mineral resources they were promised by the bill when it was signed, ultimately "threatening [their] economic livelihood."[283]

Still, there are some Alaska residents, including many from Alaska's native community who support expansion of ANILCA for its environmental and economic (tourism) benefits. In a letter submitted as testimony to the Senate's Energy and Natural Resources Committee, Julie Kitka, President of the Alaska Federation of Natives stated "ANILCA was crafted to provide subsistence priority for 'rural residents'" and that "from the statewide Native community's perspective, 'Federal overreach' is often coded language for anti-Native sentiment…used by certain urban, non-Native hunters, ranchers, and big business interests to fight Native tribes over land and resources."[283] Both labor unions and the Alaska Department of Natural Resources have also expressed their wish to protect biodiversity and access to lands for traditional activities such as subsistence hunting,

fishing, and trapping while also protecting access to adjacent non-federal lands for development opportunities.[284] It's precisely this balance between access to natural resources and protecting those natural resources from depletion or exploitation that leads to many "mixed" perspectives on expanding ANILCA (Table 20).

Alternative 3: Establish an integrated -omics research network for Arctic science

While establishing an integrated omics network generally has strong support from state and federal research-granting agencies, academics, and Democratic lawmakers, some Republican lawmakers disapprove of increased spending without a known valuation of the return on that scientific innovation. However, Senator Murkowski (R – AK) did cosponsor the introduction of the Arctic Research, Monitoring, and Observing Act of 2012 with former Senator Mark Begich (D – AK), which had a similar budget and goals, although it was not passed into law.[285] Other stakeholders like Alaska residents, labor unions, or the U.S. Navy for example, have not publicly had any strong, consistent opinions in any direction.

Based on this review of publicly-available historical and/or recent statements, a score of +1, -1.5, or 0 was assigned to each group for whether they would support, oppose, or have mixed or neutral opinions, respectively, to analyze the political feasibility of each of the proposed alternatives (Table 21). These magnitudes illustrate how a negative political viewpoint is often substantially more influential in policy decision-making.[286]

Table 21: Scores assigned to each stakeholder group for proposed policy alternatives based on political feasibility analysis

Stakeholder	UNCLOS		ANILICA		OMICS	
	Disposition	Score	Disposition	Score	Disposition	Score
Alaska residents	Support	1	Mixed	0	Neutral	0
Republicans	Mixed	0	Oppose	-1.5	Mixed	0
Democrats	Support	1	Support	1	Support	1
Labor unions	Support	1	Mixed	0	Neutral	0
Academics	Support	1	Support	1	Support	1
DOE/EPA/NSF/DOI	Support	1	Support	1	Support	1
Alaska Dept. of Natural Resources	Mixed	0	Mixed	0	Support	1
Bureau of Land Management/Fish & Wildlife	Support	1	Support	1	Support	1
Gas/petroleum/minerals industries	Support	1	Oppose	-1.5	Mixed	0
U.S. Navy/Dept. of Defense	Support	1	Neutral	0	Neutral	0
Total Score		**8**		**1**		**5**

6.4.3 *Economic feasibility*

Lastly, a critical component of the policy decision-making process is to assess the economic feasibility of each proposed alternative. For a given policy, there may be costs for multiple stakeholders including the federal government, private industry, or even individual households. Policies may require more investment from one sector over another, or investment at different times. As such, some alternatives may have more immediate benefits or result in benefits further into the future. Here, the immediate costs, investment,

or expenses directly resulting from each proposed policy alternative and the estimated future costs or benefits are summarized.

Alternative 1: Ratify UNCLOS

While there are no immediate costs or investment required to ratify UNCLOS, there have been quite a few economic analyses on the opportunity costs for not ratifying it,[279] as well as concern about the possible royalties to be paid out under Article 82 of the agreement. Article 82 states that "payments and contributions with respect to the exploitation of the continental shelf beyond 200 nautical miles" must be made at a rate of 1% of the value or volume of production at the site after the fifth year of production, increasing 1% each year after that, up to 7%. However, even though it has been estimated by the U.S. Geological Survey that 90 billion barrels of oil, 1,669 trillion cubic feet of natural gas, and 44 billion barrels of natural gas liquids exist north of the Arctic circle, these estimates are highly uncertain and refer to the entire northern area (land and sea), not all of which would be accessible and available for U.S. development.[287] This makes it impossible at this time to obtain a reliable estimate of how much may be paid out under this provision. In addition, due to the logistical and mechanical constraints on drilling and extraction in the frigid temperatures of the Arctic, it's not certain that profitable development of Arctic oil and natural gas deposits is even possible at this time either.[288] Thus, from an economic standpoint, ratifying UNCLOS is not only viable, but favorable.

Alternative 2: Increase the amount of federally-protected land under ANILCA

If this alternative were to be passed into law, the initial investment by the state and federal governments would be substantial, due to costs to create and maintain new national

parks, wildlife refuges, forests, or state parks. However, over time, federally-protected lands generally lead to increased tourism, spending, and jobs creation. It's been estimated that outdoor recreation in Alaska generates $9.5 billion in spending, supporting 92,000 jobs, and $711 million each year in state and local tax revenue.[289] Over a million out-of-state visitors engage in tourism in Alaska each year adding another ~$1.2 billion to the regional economy which includes Denali National Park, the Kenai National Wildlife Refuge, and the southeast region that includes the capital Juneau, Glacier Bay National Park and Preserve, and Chugach and Tongass National Forests. While there are challenges in estimating the economic benefits of protected lands in Alaska, economists have reduced the uncertainties in these predictions and generally conclude that additional wilderness areas would have a net-positive economic benefit in just a few years.[290, 291]

Alternative 3: Establishing an integrated -omics research network

The primary means by which an integrated research network would be created is by establishing a federal grant program to fund research related to Artic omics. Based on previous, similar programs through the National Oceanic and Atmospheric Administration and the Arctic Research, Monitoring, and Observing Act of 2012, its estimated here that the initial investment by the federal government for an effective integrated omics program would be approximately $3 million, and approximately $20 million over five years. There are no foreseeable obligatory costs to the private sector or the individual household. The economic feasibility results for each of the three alternatives has been summarized below (Table 22).

178

Table 22: Scores assigned to each policy alternative during the economic feasibility analysis for costs incurred to the federal and state government, private industry, or the individual household

	UNCLOS	**ANILCA**	**OMICS**
Federal	7	4	6
State	10	7	10
Private	2	2	10
Household	10	7	10
Total Score	**29**	**20**	**36**

6.5   Policy recommendation

Based on this feasibility analysis, each of the proposed policy alternatives had the highest score in at least one of the areas: alternative 2 (ANILCA) for technical feasibility, alternative 1 (UNCLOS) for political feasibility, and alternative 3 (OMICS) for economic feasibility. To identify which policy alternative not only had the highest feasibility, but would also address the proposed objectives, a final decision matrix incorporating all six objectives with a corresponding score was completed (Table 23). In addition to the scores, a weight was assigned to each category, demonstrating how even when a proposed policy aims to, for example, directly reduce negative impacts on the environment or minimize costs to citizens, the implementation costs taken on by the federal government and whether the bill can make it through both the House and the Senate to the President's desk, taking into account the interests of their largest supporters and various lobbying groups (political feasibility), often end up taking priority.[286]

179

Table 23: Final summary of scores for each of the objectives used to contrast each of the proposed policy alternatives

	Maximize political feasibility	Minimize costs to fed. gov.	Promote cooperation	Optimize access to nat. res.	Protect environ.	Minimize costs to AK	Total Score
Weight	**0.25**	**0.25**	**0.15**	**0.15**	**0.10**	**0.10**	
UNCLOS	8	7	10	10	5	10	**8.25**
ANILCA	1	4	6	3	8	7	**4.10**
OMICS	5	6	10	8	6	10	**7.05**

While ratifying UNCLOS (alternative 1) scored highest, losing value mainly for its neutral impact on protecting the Arctic environment, the success of this policy alternative also relies heavily on the technical feasibility of estimating energy reserves and charting the Arctic region, technology that is currently highly uncertain and relies heavily on unpredictable environmental conditions. However, creating an integrated -omics network (alternative 3) was a close second in overall score and accomplishes similar objectives. Creating a formalized network of scientists from around the world working toward a common set of deliverables would promote international cooperation, and the scientific data that would come out of these studies would inform not only fundamental biology and ecology research, but also applications in the areas of conservation and human health, key areas being impacted by climate change. The largest hurdles for creating such a research network are in the areas of technical and political feasibility. Although the data collection technologies are quite advanced, better data analysis and interpretation tools are still needed. Creating robust, agreed-upon standards and protocols for integrating -omics datasets are still in their infancy. Future research that focuses on measuring and comparing

data quality and the development of a core resource where sample preparation procedures, data acquisition and instrument parameters, and additional metadata can be stored and accessed by the community is needed. Also, a consensus needs to be developed on the ability of omics measurements to be incorporated into climate models and the amount of value added by incorporating those measurements. It could be argued however, that these are more likely to come to fruition under an organized international program like the one proposed here. Regarding political feasibility, there are many organizations that could potentially implement this policy, but as it stands, there is no clear lead agency that currently has the authority and funding to carry out the objectives associated with a program of that size.

In summary, based on the results of this analysis which compared three contrasting federal policies in terms of their technical, political, and economic feasibility to address a range of objectives for more effectively responding to climate-driven change in the Arctic, the policy that is most readily able to be implemented, will have the most positive impact on current challenges, and the least social and/or political fallout is ratifying the United Nations Law of the Sea Convention Treaty.

CHAPTER 7: OVERVIEW AND PERSPECTIVES ON EXOMETABOLOMICS IN ARCTIC SOIL

7.1   Conclusions

According to the most recent Intergovernmental Panel on Climate Change (IPCC) report, the Arctic is projected to warm an estimated $2 - 9$ °C by the year 2100.[292] One of the largest areas of uncertainty associated with these predictions is how much soil C will be lost as CO_2, CH_4, or N_2O and how to identify hotspots or hot moments of this release across the landscape—microbially-mediated processes driven by complex interactions occurring at the molecular scale that are poorly understood and/or poorly characterized in current climate models. To reduce this uncertainty around the role of organic substrate dynamics in C sequestration or release, the substrate pool must first be comprehensively characterized under a range of conditions so that individual analytes or a profile of analytes may then be identified and prioritized (sensitivity analyses), with the ultimate goal of matching these data to models across various temporal and spatial scales to reduce uncertainty in the predictions of the Arctic feedback to global climate change. To that end, the principal goals of this dissertation were to address the following two research questions:

1) Can we sensitively and robustly detect and quantify LMW DOM chemistry across space in Arctic soils using untargeted LC/MS-based exometabolomics?

2) What is the distribution of LMW DOM chemistry across a range of landscape features and conditions?

In chapters 1 and 2, I reviewed current knowledge surrounding Arctic organic matter characterization and untargeted technologies that have been developed to address this knowledge gap. Mass spectrometry-based exometabolomics has emerged as a powerful approach to sensitively and comprehensively characterize complex mixtures of

small molecules from extracellular biological matrices, including soil. The full potential of this technique had not been realized in Arctic soils however.

In chapter 3, I developed and evaluated an untargeted exometabolomics approach that employed two complementary liquid chromatography phases and two MS polarities, at the nano-scale, to expand coverage of the LMW DOM pool in Arctic soils. This optimized approach was then implemented along the length of an Arctic soil core to investigate its capabilities to sensitively and robustly provide both qualitative and relative quantitative information about the distribution of LMW DOM compounds across space (with depth). This was the first demonstration and evaluation of a nano-RP/HILIC-LC/MS approach in Arctic soil. While no single analytical approach can detect the entirety of LMW DOM compounds found in soil in an unbiased way, substantially more information was gained by combining complimentary LC/MS conditions, expanding LMW DOM coverage by 63%. Detailed molecular information, and subtle, but consistent and statistically-significant variations in the biogeochemical processing of both known and unknown compounds was provided. Of note, there were clusters of LMW DOM compounds uniquely observed at the top of the organic horizon, that were not detected deeper in the core near the transition to mineral soil, suggesting that during vertical transport, they may have become bound to trace mineral material not removed during visual inspection or that they were processed by the microbial community and released from the soil as greenhouse gases. Future targeted analyses could use this approach to monitor these compounds movement through the soil, using an isotopically labeled compound for example, to reveal additional information about whether they become bound to the soil phase, incorporated

into the microbial community, or released from the soil via mineralization or leaching, helping identify hotspots of C sinks or sources with depth.

In chapters 4 and 5, I applied the optimized approach across a range of permafrost conditions to explore trends in the LMW DOM pool with varying aboveground vegetation, topography, or level of permafrost thaw/degradation. Two sample collection techniques were employed: 1) a soil core harvest and liquid extraction and 2) a passive soil pore water collection using rhizon samples and tension lysimetry. These studies yielded new insights into the diversity of LMW DOM available to plant and microbial communities and their distribution across space in Arctic systems under a range of conditions. Across the three studies, on average, less than half of LMW DOM compounds detected were annotated by formula assignment or database searching highlighting the potential of this technique to provide information about unknowns and also one of the remaining challenges in untargeted metabolomics analyses (see below). Using the annotation data, soluble secondary organic aerosols were observed for the first time in polygonal tundra soils on the Barrow Environmental Observatory, suggesting this may be a significant organic input in these systems. More research is needed however to determine the relative importance of this process in these systems. Both polygon type and vegetation were strong predictors of LMW DOM composition, with multiple lines of evidence suggesting low-centered polygons may be hotspots of increased LMW DOM vulnerability. Because these sites are often inundated, there may exist significant methanogen populations and these studies support that LCPs may act as hotspots for CH_4 release under warmer, wetter conditions.

Finally, in chapter 6, I evaluated the potential of using multi-omics platforms to inform Arctic science from a public policy perspective and compared the creation an integrated omics network to alternative previously-proposed approaches that address various aspects of climate policy in the Arctic. Overall, this dissertation lays the analytical foundation for characterizing an information-rich signal of C vulnerability to release across space in soils, providing mechanistic insight into the controls on organic matter availability under various environmental conditions in Arctic terrestrial systems and enhancing our understanding of how this unique landscape may respond under future climate scenarios.

7.2   Remaining challenges

The advent of high-resolution mass spectrometry and bioinformatic tools has aided in providing detailed chemical information across a range of scientific disciplines, including applications in untargeted exometabolomics in soil. One of the main challenges that remains in this space is that although high-resolution mass spectrometers are becoming more common, they are still not readily-available to most geochemists or soil ecologists. In addition, there are challenges that still exist in sample preparation, data analysis, and interpretation that need to be addressed to fully realize the potential of these advanced molecular techniques.

The molecular snapshot of LMW DOM that is gained from these untargeted techniques relies heavily on consistent sample collection and processing. Soils are incredibly complex, each with varying plant inputs, microbial communities, and mineral contents that affect the proportion of LMW DOM available for biogeochemical processing. While it is known that each of these affects the accessibility of LMW DOM to the microbial

community, these processes are still not well-understood at the individual metabolite level. More research is needed to develop *in situ* microbial community measurements, to examine plant root exudate chemistry, and to study how abiotic factors (*i.e.* mineral sorption, photodegradation) influence untargeted measurements of LMW DOM. In addition, it is critical that each untargeted platform be optimized for the matrix-of-interest. Robust experimental design is essential, taking into account an appropriate replication strategy for statistical significance, randomized sampling, and the incorporation of blanks, quality controls, and internal or external standards for data validation.

Untargeted approaches also depend heavily on the data treatment procedures employed. Here, differentially-abundant features between conditions were examined after applying various filters for background, noise, and reproducibility. To ensure a conservative measurement was made and decrease instances of false-positives, the thresholds set here likely excluded "real" analytes that were undergoing biogeochemical variation between conditions. Additional data mining procedures would need to be done to explore the frequency with which those analytes occur across a variety of conditions. In this same vein, while high-mass accuracy data can be used to identify potential adducts (Chapter 4), many of those peaks also matched to a database hit suggesting data processing needs are still needed for reliable adduct identification and removal. Incorporation of fragmentation data and spectral similarity networks into data processing workflows, although in its infancy, is a promising approach to differentiate adducts from analytes (see below).[103]

As with any *-omics* technique, large datasets are generated, and many of the features reliably detected and determined to be analytes-of-interest remain unidentified. While many publicly-available databases exist, due to challenges associated with intra-lab variation between different measurement platforms, there are still many databases that are in-house or only commercially-available. Creating a core, open-access, high-mass accuracy, small molecule database would be of great value. However, details about experimental design, data quality, and reporting standards should be clearly defined and able to be evaluated by the broader metabolomics community, and methods that streamline analysis of these relationships are needed.

For both of these areas, data processing and annotation, fragmentation (MS/MS) experiments have become a powerful approach to help differentiate adducts, complexes, contaminants, knowns and unknowns. Molecular networking for example—grouping unknown (or known) compounds based on their structural similarity—that leverages high-resolution MS^2 fragmentation data is a growing area of research in the metabolomics community and should continue to be developed to assist with annotation.[103, 104, 190] Metabolite MS/MS decoy databases that assist with setting accurate and consistent false-discovery rates across platforms and datasets have recently been developed[293] but have yet to be robustly included in established data processing pipelines. In addition, in-silico fragmentation databases are predicted to grow in their importance and accuracy but should continually be validated across various metabolomic datasets.[171]

Finally, the full potential of metabolomics in Arctic soils cannot fully be realized until it has been integrated with other -omics techniques (see Chapter 6 for challenges),

and robustly scaled up to inform process-based mechanistic models of C, N, P, and/or S cycling. Future studies should include an evaluation of incorporating exometabolomics/ LMW DOM data into existing models and a sensitivity analysis on the effect of including these data on larger scale processes such as GHG emissions from Arctic soils across a range of conditions.

7.3 Future outlook

Untargeted exometabolomics in soil establishes mechanistic and stoichiometric links between soil organic matter diversity and ecosystem functioning by providing a direct measurement of the relative availability and composition of these compounds to the microbial community across space and time. Future work should include correlating shifts in LMW DOM chemistry with microbial community measurements to assist with mapping these compounds to metabolic pathways, and with other environmental variations or macroscopic landscape characteristics (*i.e.* warming, changes with seasonality, GHG emissions) to better understand to what extent GHG production potentials are linked to changes in LMW DOM chemistry. In addition, the stoichiometric ratios determined by these high-mass accuracy measurements should be correlated with ancillary measurements of pH, inorganic N or P availability, and microbial community measurements (*i.e.* nitrifiers, methanogens) to elucidate detailed profiles of the redox conditions and/or co-limitation of micronutrients to provide a metabolic footprint of microbial community function across space in soil. By coupling field measurements with incubation studies in a controlled environment, untargeted exometabolomics may also be an attractive approach to reveal compositional differences (or biases) due to defined environmental conditions (*i.e.* soil

type, microbial community structure/composition). Finally, the high-throughput nature of this technique, and the low sample volume requirement, enables this approach to also provide temporal information, such as the transformation of LMW DOM compounds over the course of a growing season or in response to stress for example. Although developed to improve our understanding of organic matter cycling in Arctic soils, the untargeted exometabolomics approach established here could feasibly be applied across a broad range of soils providing information about a variety of complex, emergent properties and processes in soil, with applications not only in climate, but agriculture, landscape management, or bioenergy as well.

REFERENCES

1.   EIA "Energy Sources have changed throughout the history of the United States". https://www.eia.gov/todayinenergy/detail.php?id=11951 (September, 2018),

2.   Scripps "The Keeling Curve". https://scripps.ucsd.edu/programs/keelingcurve/ (September, 2018),

3.   Hansen, J., Ruedy, R., Sato, M., Lo, K., Global surface temperature change. *Rev. Geophys.* **2010,** *48*, (RG4004).

4.   *National Climate Assessment*; U.S. Global Change Research Program: GlobalChange.gov, 2014.

5.   CCSP *Effects of Climate Change on Energy Production and Use in the United States*; Department of Energy, Office of Biological & Environmental Research, Washington, DC, USA: 2008.

6.   Wieder, W. R.; Bonan, G. B.; Allison, S. D., Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change* **2013,** *3*, (10), 909-912.

7.   Randall, D., Wood, R., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kottsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R., Sumi, A., Taylor, K., *Climate Models and Their Evaluation*. Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2007.

8.   Hugelius, G.; Strauss, J.; Zubrzycki, S.; Harden, J. W.; Schuur, E. A. G.; Ping, C. L.; Schirrmeister, L.; Grosse, G.; Michaelson, G. J.; Koven, C. D.; O'Donnell, J. A.; Elberling, B.; Mishra, U.; Camill, P.; Yu, Z.; Palmtag, J.; Kuhry, P., Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **2014,** *11*, (23), 6573-6593.

9.   Tarnocai, C.; Canadell, J. G.; Schuur, E. A. G.; Kuhry, P.; Mazhitova, G.; Zimov, S., Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochem. Cycles* **2009,** *23*.

10.   Richter-Menge, J., Overland, J.E., Mathis, J.T., Osborne, E. *Arctic Report Card: Update for 2017*; National Oceanic and Atmospheric Association: 2017.

11.   Abbott, B. W.; Larouche, J. R.; Jones, J. B.; Bowden, W. B.; Balser, A. W., Elevated dissolved organic carbon biodegradability from thawing and collapsing permafrost. *J. Geophys. Res. Biogeosci.* **2014,** *119*, (10), 2049-2063.

12.   Hopkins, F. M.; Torn, M. S.; Trumbore, S. E., Warming accelerates decomposition of decades-old carbon in forest soils. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109*, (26), E1753-E1761.

13.   Romanovsky, V. E.; Smith, S. L.; Christiansen, H. H., Permafrost Thermal State in the Polar Northern Hemisphere during the International Polar Year 2007-2009: a Synthesis. *Permafrost Periglac.* **2010,** *21*, (2), 106-116.

14.   Schuur, E. A. G.; McGuire, A. D.; Schadel, C.; Grosse, G.; Harden, J. W.; Hayes, D. J.; Hugelius, G.; Koven, C. D.; Kuhry, P.; Lawrence, D. M.; Natali, S. M.; Olefeldt, D.; Romanovsky, V. E.; Schaefer, K.; Turetsky, M. R.; Treat, C. C.; Vonk, J. E., Climate change and the permafrost carbon feedback. *Nature* **2015,** *520*, (7546), 171-179.

15.   Schuur, E. A. G.; Bockheim, J.; Canadell, J. G.; Euskirchen, E.; Field, C. B.; Goryachkin, S. V.; Hagemann, S.; Kuhry, P.; Lafleur, P. M.; Lee, H.; Mazhitova, G.; Nelson, F. E.; Rinke, A.; Romanovsky, V. E.; Shiklomanov, N.; Tarnocai, C.; Venevsky, S.; Vogel, J. G.; Zimov, S. A., Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *Bioscience* **2008,** *58*, (8), 701-714.

16.   Hinzman, L. D.; Deal, C. J.; McGuire, A. D.; Mernild, S. H.; Polyakov, I. V.; Walsh, J. E., Trajectory of the Arctic as an integrated system. *Ecol. Appl.* **2013,** *23*, (8), 1837-1868.

17.   Voigt, C.; Lamprecht, R. E.; Marushchak, M. E.; Lind, S. E.; Novakovskiy, A.; Aurela, M.; Martikainen, P. J.; Biasi, C., Warming of subarctic tundra increases emissions of all three important greenhouse gases - carbon dioxide, methane, and nitrous oxide. *Glob. Chang. Biol.* **2017,** *23*, (8), 3121-3138.

18.   Levy, R., Simmon, R. Heating Imbalances. https://earthobservatory.nasa.gov/Features/EnergyBalance/page3.php (September, 2018),

19.   Lynch, L. M.; Machmuller, M. B.; Cotrufo, M. F.; Paul, E. A.; Wallenstein, M. D., Tracking the fate of fresh carbon in the Arctic tundra: Will shrub expansion alter responses of soil organic matter to warming? *Soil Biol. Biochem.* **2018,** *120*, 134-144.

20.   Hobbie, S. E.; Nadelhoffer, K. J.; Hogberg, P., A synthesis: The role of nutrients as constraints on carbon balances in boreal and arctic regions. *Plant and Soil* **2002,** *242*, (1), 163-170.

21.	Wild, B.; Gentsch, N.; Capek, P.; Diakova, K.; Alves, R. J. E.; Barta, J.; Gittel, A.; Hugelius, G.; Knoltsch, A.; Kuhry, P.; Lashchinskiy, N.; Mikutta, R.; Palmtag, J.; Schleper, C.; Schnecker, J.; Shibistova, O.; Takriti, M.; Torsvik, V. L.; Urich, T.; Watzka, M.; Santruckova, H.; Guggenberger, G.; Richter, A., Plant-derived compounds stimulate the decomposition of organic matter in arctic permafrost soils. *Sci. Rep.* **2016, 6**.

22.	Vitousek, P. M.; Howarth, R. W., Nitrogen limitation on land and in the sea - how can it occur. *Biogeochemistry* **1991,** *13*, (2), 87-115.

23.	Jones, D. L.; Clode, P. L.; Kilburn, M. R.; Stockdale, E. A.; Murphy, D. V., Competition between plant and bacterial cells at the microscale regulates the dynamics of nitrogen acquisition in wheat (Triticum aestivum). *New Phytologist* **2013,** *200*, (3), 796-807.

24.	Schimel, J.; Chapin III, F. S., Tundra plant uptake of amino acid and NH4 nitrogen in situ: Plants compete well for amino acid N. *Ecology* **1996,** *77*, (7), 2142-2147.

25.	Warren, C. R., Quaternary ammonium compounds can be abundant in some soils and are taken up as intact molecules by plants. *New Phytologist* **2013,** *198*, (2), 476-85.

26.	Farrell, M.; Hill, P. W.; Farrar, J.; DeLuca, T. H.; Roberts, P.; Kielland, K.; Dahlgren, R.; Murphy, D. V.; Hobbs, P. J.; Bardgett, R. D.; Jones, D. L., Oligopeptides Represent a Preferred Source of Organic N Uptake: A Global Phenomenon? *Ecosystems* **2013,** *16*, (1), 133-145.

27.	Nashölm, T.; Kielland, K.; Ganeteg, U., Uptake of organic nitrogen by plants. *New Phytol* **2009,** *182*, (1), 31-48.

28.	Kielland, K.; McFarland, J.; Olson, K., Amino acid uptake in deciduous and coniferous taiga ecosystems. *Plant and Soil* **2006,** *288*, (1-2), 297-307.

29.	Geisseler, D.; Horwath, W. R.; Joergensen, R. G.; Ludwig, B., Pathways of nitrogen utilization by soil microorganisms - A review. *Soil Biol. Biochem.* **2010,** *42*, (12), 2058-2067.

30.	Merrick, M. J.; Edwards, R. A., Nitrogen control in bacteria. *Microbiol. Rev.* **1995,** *59*, (4), 604-&.

31.   Weintraub, M. N., Biological Phosphorus Cycling in Arctic and Alpine Soils. In *Phosphorus in Action. Soil Biology.*, Bunemann, E., Oberson, A., Frossard, E., Ed. Springer: Berlin, Heidelberg, 2011; Vol. 26.

32.   Weintraub, M. N.; Schimel, J. P., The seasonal dynamics of amino acids and other nutrients in Alaskan Arctic tundra soils. *Biogeochemistry* **2005,** *73*, (2), 359-380.

33.   Billings, S. A.; Ballantyne, F., How interactions between microbial resource demands, soil organic matter stoichiometry, and substrate reactivity determine the direction and magnitude of soil respiratory responses to warming. *Glob. Chang. Biol.* **2013,** *19*, (1), 90-102.

34.   Chen, L. Y.; Liu, L.; Mao, C.; Qin, S. Q.; Wang, J.; Liu, F. T.; Blagodatsky, S.; Yang, G. B.; Zhang, Q. W.; Zhang, D. Y.; Yu, J. C.; Yang, Y. H., Nitrogen availability regulates topsoil carbon dynamics after permafrost thaw by altering microbial metabolic efficiency. *Nature Comm.* **2018,** *9*.

35.   Chapin, F. S.; Sturm, M.; Serreze, M. C.; McFadden, J. P.; Key, J. R.; Lloyd, A. H.; McGuire, A. D.; Rupp, T. S.; Lynch, A. H.; Schimel, J. P.; Beringer, J.; Chapman, W. L.; Epstein, H. E.; Euskirchen, E. S.; Hinzman, L. D.; Jia, G.; Ping, C. L.; Tape, K. D.; Thompson, C. D. C.; Walker, D. A.; Welker, J. M., Role of land-surface changes in Arctic summer warming. *Science* **2005,** *310*, (5748), 657-660.

36.   Serreze, M. C.; Walsh, J. E.; Chapin, F. S.; Osterkamp, T.; Dyurgerov, M.; Romanovsky, V.; Oechel, W. C.; Morison, J.; Zhang, T.; Barry, R. G., Observational evidence of recent change in the northern high-latitude environment. *Climatic Change* **2000,** *46*, (1-2), 159-207.

37.   Yu, X. F.; Zou, Y. C.; Jiang, M.; Lu, X. G.; Wang, G. P., Response of soil constituents to freeze-thaw cycles in wetland soil solution. *Soil Biol. Biochem.* **2011,** *43*, (6), 1308-1320.

38.   Buckeridge, K. M.; Cen, Y. P.; Layzell, D. B.; Grogan, P., Soil biogeochemistry during the early spring in low arctic mesic tundra and the impacts of deepened snow and enhanced nitrogen availability. *Biogeochemistry* **2010,** *99*, (1-3), 127-141.

39.   Freppaz, M.; Williams, B. L.; Edwards, A. C.; Scalenghe, R.; Zanini, E., Simulating soil freeze/thaw cycles typical of winter alpine conditions: Implications for N and P availability. *Applied Soil Ecology* **2007,** *35*, (1), 247-255.

40.   Lipson, D. A.; Monson, R. K., Plant-microbe competition for soil amino acids in the alpine tundra: effects of freeze-thaw and dry-rewet events. *Oecologia* **1998,** *113*, (3), 406-414.

41.   Schimel, J. P.; Kielland, K.; Chapin, F. S., Nutrient Availability and Uptake by Tundra Plants. In *Landscape Function and Disturbance in Arctic Tundra*, Reynolds, J. F.; Tenhunen, J. D., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1996; pp 203-221.

42.   Pries, C. E. H.; Schuur, E. A. G.; Vogel, J. G.; Natali, S. M., Moisture drives surface decomposition in thawing tundra. *Journal of Geophysical Research-Biogeosciences* **2013,** *118*, (3), 1133-1143.

43.   Lara, M. J.; Nitze, I.; Grosse, G.; Martin, P.; McGuire, A. D., Reduced arctic tundra productivity linked with landform and climate change interactions. *Sci. Rep.* **2018,** *8*.

44.   Wookey, P. A.; Aerts, R.; Bardgett, R. D.; Baptist, F.; Brathen, K. A.; Cornelissen, J. H. C.; Gough, L.; Hartley, I. P.; Hopkins, D. W.; Lavorel, S.; Shaver, G. R., Ecosystem feedbacks and cascade processes: understanding their role in the responses of Arctic and alpine ecosystems to environmental change. *Glob. Chang. Biol.* **2009,** *15*, (5), 1153-1172.

45.   Olefeldt, D., Goswami, S., Grosse, G., Hayes, D., Hugelius, G., Kuhry, P., McGuire, A.D., Romanovsky, V.E., Sannel, A.B.K., Schuur, E.A.G., Turetsky, M.R., Circumpolar distribution and carbon storage of thermokarst landscapes. *Nature Comm.* **2016,** *7*.

46.   Lipson, D. A.; Zona, D.; Raab, T. K.; Bozzolo, F.; Mauritz, M.; Oechel, W. C., Water-table height and microtopography control biogeochemical cycling in an Arctic coastal tundra ecosystem. *Biogeosciences* **2012,** *9*, (1), 577-591.

47.   Zona, D.; Lipson, D. A.; Zulueta, R. C.; Oberbauer, S. F.; Oechel, W. C., Microtopographic controls on ecosystem functioning in the Arctic Coastal Plain. *Journal of Geophysical Research-Biogeosciences* **2011,** *116*.

48.   Lara, M. J.; Nitze, I.; Grosse, G.; McGuire, A. D., Tundra landform and vegetation productivity trend maps for the Arctic Coastal Plain of northern Alaska. *Scientific Data* **2018,** *5*.

49.   Iversen, C. M.; Sloan, V. L.; Sullivan, P. F.; Euskirchen, E. S.; McGuire, A. D.; Norby, R. J.; Walker, A. P.; Warren, J. M.; Wullschleger, S. D., The unseen iceberg: plant roots in arctic tundra. *New Phytologist* **2015,** *205*, (1), 34-58.

50.   Tas, N.; Prestat, E.; Wang, S.; Wu, Y. X.; Ulrich, C.; Kneafsey, T.; Tringe, S. G.; Torn, M. S.; Hubbard, S. S.; Jansson, J. K., Landscape topography structures the soil microbiome in arctic polygonal tundra. *Nature Comm.* **2018,** *9*.

51.   Hodgkins, S. B.; Tfaily, M. M.; McCalley, C. K.; Logan, T. A.; Crill, P. M.; Saleska, S. R.; Rich, V. I.; Chanton, J. P., Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, (16), 5819-5824.

52.   Sjogersten, S.; Caul, S.; Daniell, T. J.; Jurd, A. P. S.; O'Sullivan, O. S.; Stapleton, C. S.; Titman, J. J., Organic matter chemistry controls greenhouse gas emissions from permafrost peatlands. *Soil Biol. Biochem.* **2016,** *98*, 42-53.

53.   Pare, M. C.; Bedard-Haughn, A., Soil organic matter quality influences mineralization and GHG emissions in cryosols: a field-based study of sub- to high Arctic. *Glob. Chang. Biol.* **2013,** *19*, (4), 1126-1140.

54.   Xu, X. F.; Schimel, J. P.; Thornton, P. E.; Song, X.; Yuan, F. M.; Goswami, S., Substrate and environmental controls on microbial assimilation of soil organic carbon: a framework for Earth system models. *Ecol. Lett.* **2014,** *17*, (5), 547-555.

55.   Riley, W. J.; Subin, Z. M.; Lawrence, D. M.; Swenson, S. C.; Torn, M. S.; Meng, L.; Mahowald, N. M.; Hess, P., Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences* **2011,** *8*, (7), 1925-1953.

56.   Tang, J. Y.; Riley, W. J., Weaker soil carbon-climate feedbacks resulting from microbial and abiotic interactions. *Nat. Clim. Change* **2015,** *5*, (1), 56-60.

57.   Lehmann, J.; Kleber, M., The contentious nature of soil organic matter. *Nature* **2015,** *528*, (7580), 60-68.

58.   Six, J.; Jastrow, J., *Organic Matter Turnover*. 2002.

59.   Drake, T. W.; Wickland, K. P.; Spencer, R. G. M.; McKnight, D. M.; Striegl, R. G., Ancient low-molecular-weight organic acids in permafrost fuel rapid carbon dioxide production upon thaw. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, (45), 13946-13951.

60.   van Hees, P. A. W.; Jones, D. L.; Finlay, R.; Godbold, D. L.; Lundstomd, U. S., The carbon we do not see - the impact of low molecular weight compounds on carbon dynamics and respiration in forest soils: a review. *Soil Biol. Biochem.* **2005,** *37*, (1), 1-13.

61.   Raynaud, X., Nunan, N., Spatial ecology of bacteria at the microscale in soil. *PLoS One* **2014,** *9*, (1), e87217.

62.   Boddy, E.; Hill, P. W.; Farrar, J.; Jones, D. L., Fast turnover of low molecular weight components of the dissolved organic carbon pool of temperate grassland field soils. *Soil Biol. Biochem.* **2007,** *39*, (4), 827-835.

63.   Chen, H. M.; Yang, Z. M.; Chu, R. K.; Tolic, N.; Liang, L. Y.; Graham, D. E.; Wullschleger, S. D.; Gu, B. H., Molecular Insights into Arctic Soil Organic Matter Degradation under Warming. *Environ. Sci. Technol.* **2018,** *52*, (8), 4555-4564.

64.   Yang, Z.; Wullschleger, S. D.; Liang, L.; Graham, D. E.; Gu, B., Effects of warming on the degradation and production of low-molecular-weight labile organic carbon in an Arctic tundra soil. *Soil Biol. Biochem.* **2016,** *95*, 202-211.

65.   Boddy, E.; Roberts, P.; Hill, P. W.; Farrar, J.; Jones, D. L., Turnover of low molecular weight dissolved organic C (DOC) and microbial C exhibit different temperature sensitivities in Arctic tundra soils. *Soil Biol. Biochem.* **2008,** *40*, (7), 1557-1566.

66.   Herndon, E. M.; Yang, Z. M.; Bargar, J.; Janot, N.; Regier, T.; Graham, D.; Wullschleger, S.; Gu, B. H.; Liang, L. Y., Geochemical drivers of organic matter decomposition in arctic tundra soils. *Biogeochemistry* **2015,** *126*, (3), 397-414.

67.   Judd, K. E.; Crump, B. C.; Kling, G. W., Variation in dissolved organic matter controls bacterial production and community composition. *Ecology* **2006,** *87*, (8), 2068-2079.

68.   Guigue, J.; Leveque, J.; Mathieu, O.; Schmitt-Kopplin, P.; Lucio, M.; Arrouays, D.; Jolivet, C.; Dequiedt, S.; Prevost-Boure, N. C.; Ranjard, L., Water-extractable organic

matter linked to soil physico-chemistry and microbiology, at the regional scale. *Soil Biol. Biochem.* **2015,** *84*, 158-167.

69.   Warren, C. R., Comparison of methods for extraction of organic N monomers from soil microbial biomass. *Soil Biol. Biochem.* **2015,** *81*, 67-76.

70.   Beisken, S.; Eiden, M.; Salek, R. M., Getting the right answers: understanding metabolomics challenges. *Expert Rev. Mol. Diagn.* **2015,** *15*, (1), 97-109.

71.   Darrouzet-Nardi, A.; Ladd, M. P.; Weintraub, M. N., Fluorescent microplate analysis of amino acids and other primary amines in soils. *Soil Biol. Biochem.* **2013,** *57*, 78-82.

72.   Foster, A.; Jones, D. L.; Cooper, E. J.; Roberts, P., Freeze-thaw cycles have minimal effect on the mineralisation of low molecular weight, dissolved organic carbon in Arctic soils. *Polar Biol.* **2016,** *39*, (12), 2387-2401.

73.   Jones, D. L.; Kielland, K.; Sinclair, F. L.; Dahlgren, R. A.; Newsham, K. K.; Farrar, J. F.; Murphy, D. V., Soil organic nitrogen mineralization across a global latitudinal gradient. *Global Biogeochem. Cycles* **2009,** *23*.

74.   Kiersch, K.; Kruse, J.; Regier, T. Z.; Leinweber, P., Temperature resolved alteration of soil organic matter composition during laboratory heating as revealed by C and N XANES spectroscopy and Py-FIMS. *Thermochim. Acta* **2012,** *537*, 36-43.

75.   Jones, D. L.; Willett, V. B., Experimental evaluation of methods to quantify dissolved organic nitrogen (DON) and dissolved organic carbon (DOC) in soil. *Soil Biol. Biochem.* **2006,** *38*, (5), 991-999.

76.   Gardeñas, A. I.; Agren, G. I.; Bird, J. A.; Clarholm, M.; Hallin, S.; Ineson, P.; Katterer, T.; Knicker, H.; Nilsson, S. I.; Nashölm, T.; Ogle, S.; Paustian, K.; Persson, T.; Stendahl, J., Knowledge gaps in soil carbon and nitrogen interactions - From molecular to global scale. *Soil Biol. Biochem.* **2011,** *43*, (4), 702-717.

77.   Grewer, D. M.; Lafreniere, M. J.; Lamoureux, S. F.; Simpson, M. J., Redistribution of soil organic matter by permafrost disturbance in the Canadian High Arctic. *Biogeochemistry* **2016,** *128*, (3), 397-415.

78.   Mann, P. J.; Davydova, A.; Zimov, N.; Spencer, R. G. M.; Davydov, S.; Bulygina, E.; Zimov, S.; Holmes, R. M., Controls on the composition and lability of

dissolved organic matter in Siberia's Kolyma River basin. *J. Geophys. Res. Biogeosci.* **2012,** *117*.

79.   White, D. M.; Garland, D. S.; Ping, C. L.; Michaelson, G., Characterizing soil organic matter quality in arctic soil by cover type and depth. *Cold Reg. Sci. Technol.* **2004,** *38*, (1), 63-73.

80.   Pautler, B. G.; Austin, J.; Otto, A.; Stewart, K.; Lamoureux, S. F.; Simpson, M. J., Biomarker assessment of organic matter sources and degradation in Canadian High Arctic littoral sediments. *Biogeochemistry* **2010,** *100*, (1-3), 75-87.

81.   Mann, B. F.; Chen, H. M.; Herndon, E. M.; Chu, R. K.; Tolic, N.; Portier, E. F.; Chowdhury, T. R.; Robinson, E. W.; Callister, S. J.; Wullschleger, S. D.; Graham, D. E.; Liang, L. Y.; Gu, B. H., Indexing Permafrost Soil Organic Matter Degradation Using High-Resolution Mass Spectrometry. *PLoS One* **2015,** *10*, (6).

82.   Choi, J. H.; Kim, Y. G.; Lee, Y. K.; Pack, S. P.; Jung, J. Y.; Jang, K. S., Chemical characterization of dissolved organic matter in moist acidic tussock tundra soil using ultra-high resolution 15T FT-ICR mass spectrometry. *Biotechnol. Bioprocess Eng.* **2017,** *22*, (5), 637-646.

83.   Sleighter, R. L.; Hatcher, P. G., The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **2007,** *42*, (5), 559-574.

84.   Warren, C. R., Development of liquid chromatography mass spectrometry method for analysis of organic N monomers in soil. *Soil Biol. Biochem.* **2014,** *78*, 233-242.

85.   Jenkins, S.; Swenson, T. L.; Lau, R.; Rocha, A. M.; Aaring, A.; Hazen, T. C.; Chakraborty, R.; Northen, T. R., Construction of Viable Soil Defined Media Using Quantitative Metabolomics Analysis of Soil Metabolites. *Front. Microbiol.* **2017,** *8*.

86.   Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000,** *18*, (11), 1157-1161.

87.   Silva, L. P.; Northen, T. R., Exometabolomics and MSI: deconstructing how cells interact to transform their small molecule environment. *Curr. Opin. Biotechnol.* **2015,** *34*, 209-216.

88.	Lubbe, A., Northen, T., Exometabolomics for Linking Soil Carbon Dynamics to Microbial Communities. In *Microbial Metabolomics*, Beale, D. J., Ed. Springer International Publishing: Switzerland, 2016; pp 119-145.

89.	Swenson, T. L.; Karaoz, U.; Swenson, J. M.; Bowen, B. P.; Northen, T. R., Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nature Comm.* **2018,** *9*.

90.	Jones, O. A. H., Stepanian, S., Lofts, S., Svendsen, C., Spurgeon, D.J. Maguire, M.L., Griffin, J.L., Metabolomic analysis of soil communities can be used for pullution assessment. *Environ. Toxicol. Chem.* **2014,** *33*, (61-64).

91.	Grivet, J. P., Delort, A.M. Portais, J.C., NMR and microbiology: from physiology to metabolomics. *Biochimie Metabolic NMR* **2013,** *85*, 823-840.

92.	Viant, M. R.; Sommer, U., Mass spectrometry based environmental metabolomics: a primer and review. *Metabolomics* **2013,** *9*, (1), S144-S158.

93.	Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O., FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Anal. Chem.* **2009,** *81*, (24), 10038-10048.

94.	Swenson, T. L.; Bowen, B. P.; Nico, P. S.; Northen, T. R., Competitive sorption of microbial metabolites on an iron oxide mineral. *Soil Biol. Biochem.* **2015,** *90*, 34-41.

95.	Kakumanu, M. L.; Cantrell, C. L.; Williams, M. A., Microbial community response to varying magnitudes of desiccation in soil: A test of the osmolyte accumulation hypothesis. *Soil Biol. Biochem.* **2013,** *57*, 644-653.

96.	Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D., Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* **2014,** *87*, 12-25.

97.	Vorkas, P. A.; Isaac, G.; Anwar, M. A.; Davies, A. H.; Want, E. J.; Nicholson, J. K.; Holmes, E., Untargeted UPLC-MS Profiling Pipeline to Expand Tissue Metabolome Coverage: Application to Cardiovascular Disease. *Anal. Chem.* **2015,** *87*, (8), 4184-4193.

98.	Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W., LC-MS-based metabolomics. *Mol. Biosyst.* **2012,** *8*, (2), 470-481.

99.   Evans, A. M.; DeHaven, C. D.; Barrett, T.; Mitchell, M.; Milgram, E., Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems. *Anal. Chem.* **2009,** *81*, (16), 6656-6667.

100.   Baran, R., Brodie, E.L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U.N., Chakraborty, R., Bowen, B.P., Karaoz, U., Cadillo-Quiroz, H., Garcia-Pichel, F., Northen, T.R., Exometabolite niche partitioning among sympatric soil bacteria. *Nature Comm.* **2015,** *6*, (8289).

101.   Ding, J. H.; Wang, X. X.; Zhang, T. L.; Li, Q. M.; Luo, M. B., Optimization of RP-HPLC analysis of low molecular weight organic acids in soil. *Journal of Liquid Chromatography & Related Technologies* **2006,** *29*, (1), 99-112.

102.   Ramautar, R., Somsen, G., de Jong, G. J., CE-MS for metabolomics: Developments and applications in the period 2014-2016. *Electrophoresis* **2017,** *38*, (1), 190-202.

103.   Sedio, B. E.; Boya, C. A.; Rojas Echeverri, J. C., A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. *Appl. Plant Sci.* **2018,** *6*, (3).

104.   Hufsky, F.; Bocker, S., Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrom. Rev.* **2017,** *36*, (5), 624-633.

105.   Morreel, K.; Kim, H.; Lu, F. C.; Dima, O.; Akiyama, T.; Vanholme, R.; Niculaes, C.; Goeminne, G.; Inze, D.; Messens, E.; Ralph, J.; Boerjan, W., Mass Spectrometry-Based Fragmentation as an Identification Tool in Lignomics. *Anal. Chem.* **2010,** *82*, (19), 8095-8105.

106.   Kind, T. F., O., Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalyt. Rev.* **2010,** *2*, 23-60.

107.   Zhang, R.; Watson, D. G.; Wang, L. J.; Westrop, G. D.; Coombs, G. H.; Zhang, T., Evaluation of mobile phase characteristics on three zwitterionic columns in hydrophilic interaction liquid chromatography mode for liquid chromatography-high

resolution mass spectrometry based untargeted metabolite profiling of Leishmania parasites. *J. Chromatogr. A* **2014,** *1362*, 168-179.

108.   Guo, Y.; Gaiki, S., Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. *J. Chromatogr. A* **2005,** *1074*, (1-2), 71-80.

109.   Buszewski, B.; Noga, S., Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. *Anal. Bioanal. Chem.* **2012,** *402*, (1), 231-247.

110.   Chirita, R., West, C., Zubrzycki, S., Finaru, A., Elfakir, C., Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography. *J. Chromatogr. A* **2011,** *1218*, 5939-5963.

111.   Contrepois, K.; Jiang, L. H.; Snyder, M., Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol. Cell. Proteomics* **2015,** *14*, (6), 1684-1695.

112.   Ivanisevic, J.; Zhu, Z. J.; Plate, L.; Tautenhahn, R.; Chen, S.; O'Brien, P. J.; Johnson, C. H.; Marletta, M. A.; Patti, G. J.; Siuzdak, G., Toward 'Omic Scale Metabolite Profiling: A Dual Separation-Mass Spectrometry Approach for Coverage of Lipid and Central Carbon Metabolism. *Anal. Chem.* **2013,** *85*, (14), 6876-6884.

113.   Gray, N.; Heaton, J.; Musenga, A.; Cowan, D. A.; Plumb, R. S.; Smith, N. W., Comparison of reversed-phase and hydrophilic interaction liquid chromatography for the quantification of ephedrines using medium-resolution accurate mass spectrometry. *J. Chromatogr. A* **2013,** *1289*, 37-46.

114.   Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G., Evaluation of Coupling Reversed Phase, Aqueous Normal Phase, and Hydrophilic Interaction Liquid Chromatography with Orbitrap Mass Spectrometry for Metabolomic Studies of Human Urine. *Anal. Chem.* **2012,** *84*, (4), 1994-2001.

115.   Qin, F.; Zhao, Y. Y.; Sawyer, M. B.; Li, X. F., Column-switching reversed phase-hydrophilic interaction liquid chromatography/tandem mass spectrometry method for determination of free estrogens and their conjugates in river water. *Anal. Chim. Acta* **2008,** *627*, (1), 91-98.

116.	Hutta, M., Chalanyova, M., Halko, R., Gora, R., Dokupilova, S., Rybar, I., Reversed phase liquid chromatography trace analysis of pesticides in soil by on-column sample pumping large volume injection and UV detection. *J. Sep. Sci.* **2009,** *32*, (12), 2034-42.

117.	Wells, M. J. M., Michael, J.L., Neary, D.J., Determination of picloram in soil and water by reversed-phase liquid chromatography. *Arch. Environ. Contam. Toxicol.* **1984,** *13*, (2), 231-235.

118.	Tian, Q., Zhou, Z., Lv, C., Huang, Y., Ren, L., Simultaneous determination of paclobutrazol and myclobutanil enantiomers in water and soil using enantioselective reversed-phase liquid chromatography. *Anal. Methods* **2010,** *2*, 617-622.

119.	Trubetskoi, O. A., Trubetskaya, O.E., Reversed-phase high-performance liquid chromatography of the stabel electrophoretic fractions of soil humic acids. *Eurasian Soil Science* **2015,** *48*, (2), 148-156.

120.	Gao, J. J.; Helmus, R.; Cerli, C.; Jansen, B.; Wang, X.; Kalbitz, K., Robust analysis of underivatized free amino acids in soil by hydrophilic interaction liquid chromatography coupled with electrospray tandem mass spectrometry. *J. Chromatogr. A* **2016,** *1449*, 78-88.

121.	Olofsson, M. A.; Bylund, D., Liquid Chromatography with Electrospray Ionization and Tandem Mass Spectrometry Applied in the Quantitative Analysis of Chitin-Derived Glucosamine for a Rapid Estimation of Fungal Biomass in Soil. *Int. J. Anal. Chem.* **2016**.

122.	Sandron, S., Rojas, A., Wilson, R., Davies, N., Haddad, P., Shellie, R., Nesterenko, P., Kelleher, B., Paull, B., Chromatographic methods for the isolation, separation and characterisation of dissolved organic matter. *Environ Sci: Processes Impacts* **2015,** *17*, 1531-1567.

123.	Nguyen, H. P.; Schug, K. A., The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications. *J. Sep. Sci.* **2008,** *31*, (9), 1465-1480.

124.	Yamashita, M., Fenn, J.B., Electrospray ion-source-another variation on the free-jet theme. *J. Phys. Chem.* **1984,** *88*, 4451-9.

125.   Dole, M., Cox, H.L, Gieniec, J., Electrospray mass-spectroscopy. *Adv. Chem. Ser.* **1973**, 73-84.

126.   Banerjee, S.; Mazumdar, S., Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**.

127.   Cole, R. B., *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications*. John Wiley and Sons: New York, 1997.

128.   Juraschek, R.; Dulcks, T.; Karas, M., Nanoelectrospray - More than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **1999,** *10*, (4), 300-308.

129.   Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **1996,** *68*, (1), 1-8.

130.   Edwards, J. L.; Edwards, R. L.; Reid, K. R.; Kennedy, R. T., Effect of decreasing column inner diameter and use of off-line two-dimensional chromatography on metabolite detection in complex mixtures. *J. Chromatogr. A* **2007,** *1172*, (2), 127-134.

131.   Li, Z. A.; Zou, B.; Xia, H. P.; Ding, Y. Z.; Tan, W. N.; Fu, S. L., Role of low-molecule-weight organic acids and their salts in regulating soil pH(-1). *Pedosphere* **2008,** *18*, (2), 137-148.

132.   Kandiah, M.; Urban, P. L., Advances in ultrasensitive mass spectrometry of organic molecules. *Chem. Soc. Rev.* **2013,** *42*, (12), 5299-5322.

133.   Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007,** *26*, (1), 51-78.

134.   Tfaily, M. M.; Chu, R. K.; Tolic, N.; Roscioli, K. M.; Anderton, C. R.; Pasa-Tolic, L.; Robinson, E. W.; Hess, N. J., Advanced Solvent Based Methods for Molecular Characterization of Soil Organic Matter by High-Resolution Mass Spectrometry. *Anal. Chem.* **2015,** *87*, (10), 5206-5215.

135.   Ohno, T.; Parr, T. B.; Gruselle, M. C. I.; Fernandez, I. J.; Sleighter, R. L.; Hatcher, P. G., Molecular Composition and Biodegradability of Soil Organic Matter: A Case Study Comparing Two New England Forest Types. *Environ. Sci. Technol.* **2014,** *48*, (13), 7229-7236.

136.   Reemtsma, T., Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry Status and needs. *J. Chromatogr. A* **2009,** *1216*, (18), 3687-3701.

137.   Dunn, W. B., Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian, and plant metabolomes. *Phys. Biol.* **2008,** *5*, (1), 24.

138.   Cajka, T.; Fiehn, O., Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **2016,** *88*, (1), 524-545.

139.   ORNL NGEE Arctic: Next-Generation Ecosystem Experiments http://ngee-arctic.ornl.gov/ (11/10/2016),

140.   Darrouzet-Nardi, A.; Weintraub, M. N., Evidence for spatially inaccessible labile N from a comparison of soil core extractions and soil pore water lysimetry. *Soil Biol. Biochem.* **2014,** *73*, 22-32.

141.   Guigue, J., Mathieu, O., Leveque, J., Mounier, S., Laffont, R., Maron, P.A., Navarro, N., Chateau, C., Amiotte-Suchet, P., Lucas, Y., A comparison of extraction procedures for water-extractable organic matter in soils. *Eur. J. Soil Sci.* **2014,** *65*, 520-530.

142.   Swenson, T. L.; Jenkins, S.; Bowen, B. P.; Northen, T. R., Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.* **2015,** *80*, 189-198.

143.   Wang, X. X.; Li, Q. M.; Ding, J. H.; Luo, M. B.; Zhang, T. L.; Zhou, Y. Y., An improved method for the extraction of low molecular weight organic acids in variable charge soils. *Anal. Sci.* **2007,** *23*, (5), 539-543.

144.   Ladd, M. P., Wullschleger, S.D., Hettich, R. L. In *Development and evaluation of a dual-separation, high-resolution, nano-ESI-LC-MS/MS approach for dissolved soil organic matter characterization*, 64th Annual ASMS Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, 2016; American Society for Mass Spectrometry: San Antonio, TX, 2016.

145.   Payne, J. M.; Smith, M. W., Peptide transport by microorganisms. *Adv. Microb. Physiol.* **1994,** *36*, 1-80.

146.   Ladd, M. P., Abraham, P., Giannone, R., Norby, R., Hettich, R.L. In *Characterizing the range of low molecular weight organic compounds in nitrogen-limited Arctic soils using electrospray mass spectrometry*, 63rd Annual ASMS Conference on Mass Spectrometry and Allied Topics, St. Louis, MO, 2015; St. Louis, MO, 2015.

147.   Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W., Potential of metabolomics as a functional genomics tool. *Trends in Plant Sci.* **2004,** *9*, (9), 418-425.

148.   Li, B.; Tang, J.; Yang, Q. X.; Li, S.; Cui, X. J.; Li, Y. H.; Chen, Y. Z.; Xue, W. W.; Li, X. F.; Zhu, F., NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* **2017,** *45*, (W1), W162-W170.

149.   Berg, M.; Vanaerschot, M.; Jankevics, A.; Cuypers, B.; Breitling, R.; Dujardin, J.-C., LC-MS metabolomics from study design to data-analysis – using a versatile pathogen as a test case. *Comput. Struct. Biotechnol. J.* **2013,** *4*, e201301002.

150.   Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, H. C., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011,** *6*, (7), 1060-1083.

151.   Naidong, W.; Shou, W.; Chen, Y. L.; Jiang, X. Y., Novel liquid chromatographic-tandem mass spectrometric methods using silica columns and aqueous-organic mobile phases for quantitative analysis of polar ionic analytes in biological fluids. *J. Chromatogr. B* **2001,** *754*, (2), 387-399.

152.   Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D., MZmine 2 Data-Preprocessing To Enhance Molecular Networking Reliability. *Anal. Chem.* **2017,** *89*, (15), 7836-7840.

153.   Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G., Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: Quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal. Chem.* **2006,** *78*, (10), 3289-3295.

154. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*.

155. De Livera, A. M.; Dias, D. A.; De Souza, D.; Rupasinghe, T.; Pyke, J.; Tull, D.; Roessner, U.; McConville, M.; Speed, T. P., Normalizing and Integrating Metabolomics Data. *Anal. Chem.* **2012,** *84*, (24), 10768-10776.

156. De Livera, A. M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P., Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal. Chem.* **2015,** *87*, (7), 3606-3615.

157. Polpitiya, A. D. Q., W.J.; Jaitly, N.; Petyuk, V.A.; Adkins, J.N.; Camp, D.G. 2nd; Anderson, G. A. S., R.D., DanTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **2008,** *24*, (13), 1556-1558.

158. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **2016,** *13*, (9), 731-740.

159. SAS-Institute, JMP Software. **1989-2007**.

160. Kalivodová, A.; Hron, K.; Filzmoser, P.; Najdekr, L.; Janečková, H.; Adam, T., PLS-DA for compositional data with application to metabolomics. *J. Chemom.* **2015,** *29*, (1), 21-28.

161. Barnes, S.; Benton, H. P.; Casazza, K.; Cooper, S. J.; Cui, X.; Du, X.; Engler, J.; Kabarowski, J. H.; Li, S.; Pathmasiri, W.; Prasain, J. K.; Renfrow, M. B.; Tiwari, H. K., Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. *J. Mass Spectrom.* **2016,** *51*, (8), 535-548.

162. Jones, E., Oliphant, T., Peterson, P. , {SciPy}: Open source scientific tools for {Python}. **2014**.

163. Zhou B, W. J., Ressom HW, MetaboSearch: Tool for mass-based metabolite identification using multiple databases. *PLoS One* **2012,** *7*, (6).

164.   Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K., KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017,** *45*, (D1), D353-d361.

165.   Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27*, (6), 747-751.

166.   Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.* **2008,** *26*, (2), 162-164.

167.   Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L. Y.; He, J. E.; He, S. Q.; Shoemaker, B. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H., PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016,** *44*, (D1), D1202-D1213.

168.   Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013,** *41*, (D1), D801-D807.

169.   Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S., LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **2007,** *35*, W606-W612.

170.   (PMN), Plant Metabolic Network: PlantCyc Database - www.plantcyc.org. **2018**.

171.   Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z. J.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O., Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2018,** *37*, (4), 513-532.

172.   Koch, B. P.; Dittmar, T.; Witt, M.; Kattner, G., Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **2007,** *79*, (4), 1758-1763.

209

173.   Kind, T.; Fiehn, O., Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007,** *8*.

174.   Kujawinski, E. B.; Behn, M. D., Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **2006,** *78*, (13), 4363-4373.

175.   Tfaily, M. M.; Wilson, R. M.; Cooper, W. T.; Kostka, J. E.; Hanson, P.; Chanton, J. P., Vertical Stratification of Peat Pore Water Dissolved Organic Matter Composition in a Peat Bog in Northern Minnesota. *Journal of Geophysical Research-Biogeosciences* **2018,** *123*, (2), 479-494.

176.   Patti, G. J., Separation strategies for untargeted metabolomics. *J. Sep. Sci.* **2011,** *34*, (24), 3460-3469.

177.   Fairchild, J. N.; Horvath, K.; Gooding, J. R.; Campagna, S. R.; Guiochon, G., Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. *J. Chromatogr. A* **2010,** *1217*, (52), 8161-8166.

178.   Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G., Expanding Coverage of the Metabolome for Global Metabolite Profiling. *Anal. Chem.* **2011,** *83*, (6), 2152-2161.

179.   Spagou, K.; Tsoukali, H.; Raikos, N.; Gika, H.; Wilson, I. D.; Theodoridis, G., Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *J. Sep. Sci.* **2010,** *33*, (6-7), 716-727.

180.   Bajad, S. U.; Lu, W. Y.; Kimball, E. H.; Yuan, J.; Peterson, C.; Rabinowitz, J. D., Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **2006,** *1125*, (1), 76-88.

181.   Creydt, M.; Fischer, M., Plant Metabolomics: Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Nontargeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Anal. Chem.* **2017,** *89*, (19), 10474-10486.

182.   Simmons, K., Deatrick, J., Lewis, B., Operating Procedure: Soil Sampling. In Science and Ecosystem Support Division, U.S. EPA: 2014.

183.   Schmidt, M. W.; Torn, M. S.; Abiven, S.; Dittmar, T.; Guggenberger, G.; Janssens, I. A.; Kleber, M.; Kogel-Knabner, I.; Lehmann, J.; Manning, D. A.; Nannipieri,

P.; Rasse, D. P.; Weiner, S.; Trumbore, S. E., Persistence of soil organic matter as an ecosystem property. *Nature* **2011,** *478*, (7367), 49-56.

184.   Teleki, A.; Sanchez-Kopper, A.; Takors, R., Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry. *Anal. Biochem.* **2015,** *475*, 4-13.

185.   Kostiainen, R.; Kauppila, T. J., Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2009,** *1216*, (4), 685-699.

186.   Williams, M. A.; Xia, K., Characterization of the water soluble soil organic pool following the rewetting of dry soil in a drought-prone tallgrass prairie. *Soil Biol. Biochem.* **2009,** *41*, (1), 21-28.

187.   Hayat, S.; Hayat, Q.; Alyemeni, M. N.; Wani, A. S.; Pichtel, J.; Ahmad, A., Role of proline under changing environments: a review. *Plant signaling & behavior* **2012,** *7*, (11), 1456-1466.

188.   Gargallo-Garriga, A.; Preece, C.; Sardans, J.; Oravec, M.; Urban, O.; Peñuelas, J., Root exudate metabolomes change under drought and show limited capacity for recovery. *Sci. Rep.* **2018,** *8*, (1), 12696.

189.   Rady, M. M.; Taha, R. S.; Mahdi, A. H. A., Proline enhances growth, productivity and anatomy of two varieties of *Lupinus termis* L. grown under salt stress. *South African Journal of Botany* **2016,** *102*, 221-227.

190.   de Oliveira, G. G.; Neto, F. C.; Demarque, D. P.; Pereira, J. A. D.; Peixoto, R. C. S.; de Melo, S. J.; Almeida, J.; Lopes, J. L. C.; Lopes, N. P., Dereplication of Flavonoid Glycoconjugates from Adenocalymma imperatoris-maximilianii by Untargeted Tandem Mass Spectrometry-Based Molecular Networking. *Planta Med.* **2017,** *83*, (7), 636-646.

191.   Liljedahl, A. K.; Boike, J.; Daanen, R. P.; Fedorov, A. N.; Frost, G. V.; Grosse, G.; Hinzman, L. D.; Iijma, Y.; Jorgenson, J. C.; Matveyeva, N.; Necsoiu, M.; Raynolds, M. K.; Romanovsky, V. E.; Schulla, J.; Tape, K. D.; Walker, D. A.; Wilson, C. J.; Yabuki, H.; Zona, D., Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology. *Nature Geoscience* **2016,** *9*, (4), 312-+.

192.   Wainwright, H. M.; Dafflon, B.; Smith, L. J.; Hahn, M. S.; Curtis, J. B.; Wu, Y. X.; Ulrich, C.; Peterson, J. E.; Torn, M. S.; Hubbard, S. S., Identifying multiscale

zonation and assessing the relative importance of polygon geomorphology on carbon fluxes in an Arctic tundra ecosystem. *Journal of Geophysical Research-Biogeosciences* **2015,** *120*, (4), 788-808.

193.   Hubbard, S. S.; Gangodagamage, C.; Dafflon, B.; Wainwright, H.; Peterson, J.; Gusmeroli, A.; Ulrich, C.; Wu, Y.; Wilson, C.; Rowland, J.; Tweedie, C.; Wullschleger, S. D., Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets. *Hydrogeology Journal* **2013,** *21*, (1), 149-169.

194.   Norby, R. J.; Sloan, V. L.; Iversen, C. M.; Childs, J., Controls on Fine-Scale Spatial and Temporal Variability of Plant-Available Inorganic Nitrogen in a Polygonal Tundra Landscape. *Ecosystems* **2018**.

195.   Newman, B. D.; Throckmorton, H. M.; Graham, D. E.; Gu, B.; Hubbard, S. S.; Liang, L.; Wu, Y.; Heikoop, J. M.; Herndon, E. M.; Phelps, T. J.; Wilson, C. J.; Wullschleger, S. D., Microtopographic and depth controls on active layer chemistry in Arctic polygonal ground. *Geophys. Res. Lett.* **2015,** *42*, (6), 1808-1817.

196.   Sloan, V. L., Liebig, J., Hahn, M., Curtis, B., Brooks, J., Rogers, A., Iversen, C., Norby, R. , Soil temperature, soil moisture, and thaw depth, Barrow, Alaska Ver. 1. In Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, TN, USA: Next Generation Ecosystem Experiments Arcic Data Collection, 2014.

197.   NOAA National Overview. http://www.ncdc.noaa.gov/sotc/national/2014/5 (August 12, 2014),

198.   Chowdhury, T. R.; Herndon, E. M.; Phelps, T. J.; Elias, D. A.; Gu, B. H.; Liang, L. Y.; Wullschleger, S. D.; Graham, D. E., Stoichiometry and temperature sensitivity of methanogenesis and CO2 production from saturated polygonal tundra in Barrow, Alaska. *Glob. Chang. Biol.* **2015,** *21*, (2), 722-737.

199.   Kaiser, K.; Kalbitz, K., Cycling downwards - dissolved organic matter in soils. *Soil Biol. Biochem.* **2012,** *52*, 29-32.

200.   Strom, L.; Tagesson, T.; Mastepanov, M.; Christensen, T. R., Presence of Eriophorum scheuchzeri enhances substrate availability and methane emission in an Arctic wetland. *Soil Biol. Biochem.* **2012,** *45*, 61-70.

201.   Kielland, K., *Short-circuiting the nitrogen cycle: Ecophysiological strategies of nitrogen uptake in plants from marginal environments*. 2001; p 376-398.

202.   Verchot, L. V.; Dutaur, L.; Shepherd, K. D.; Albrecht, A., Organic matter stabilization in soil aggregates: Understanding the biogeochemical mechanisms that determine the fate of carbon inputs in soils. *Geoderma* **2011,** *161*, (3-4), 182-193.

203.   Kujawinski, E. B., Freitas, M.A., Zang, X., Hatcher, P.G., Greenchurch, K.B., Jones, R.B., The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. *Org. Geochem.* **2002,** *33*, 171.

204.   Brown, T. L., Rice, J.A., Effect of experimental parameters on the ESI FT-ICR mass spectrum of fulvic acid. *Anal. Chem.* **2000,** *72*, 384.

205.   Lipson, D. A.; Raab, T. K.; Goria, D.; Zlamal, J., The contribution of Fe(III) and humic acid reduction to ecosystem respiration in drained thaw lake basins of the Arctic Coastal Plain. *Global Biogeochem. Cycles* **2013,** *27*, (2), 399-409.

206.   Lipson, D. A.; Jha, M.; Raab, T. K.; Oechel, W. C., Reduction of iron (III) and humic substances plays a major role in anaerobic respiration in an Arctic peat soil. *Journal of Geophysical Research-Biogeosciences* **2010,** *115*.

207.   Kim, S.; Kramer, R. W.; Hatcher, P. G., Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **2003,** *75*, (20), 5336-5344.

208.   Antony, R., Grannas, A., Willoughby, A., Sleighter, R., Thamban, M., Hatcher, P., Origin and sources of dissolved organic matter in snow on the East Antarctic ice sheet. *Environ. Sci. Technol.* **2014,** *48*, 6151-6159.

209.   Mendez-Millan, M., Dignac, M.F., Rumpel, C., Rasse, D.P., Derenne, S., Molecular dynamics of shoot vs. root biomarkers in an agricultural soil estimated by natural abundance 13C labeling. *Soil Biol. Biochem.* **2010,** *42*, 169-177.

210.   Rivas-Ubach, A.; Liu, Y.; Bianchi, T. S.; Tolic, N.; Jansson, C.; Pasa-Tolic, L., Moving beyond the van Krevelen Diagram: A New Stoichiometric Approach for Compound Classification in Organisms. *Anal. Chem.* **2018,** *90*, (10), 6152-6160.

211.   Kuzyakov, Y., Review: Factors affecting rhizosphere priming. *Journal of Plant Nutrition and Soil Science* **2002,** *165*, (4), 382-396.

212. Reemtsma, T.; These, A.; Venkatachari, P.; Xia, X. Y.; Hopke, P. K.; Springer, A.; Linscheid, M., Identification of fulvic acids and sulfated and nitrated analogues in atmospheric aerosol by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2006,** *78*, (24), 8299-8304.

213. EPA Secondary Organic Aerosol (SOAs) Research. https://www.epa.gov/air-research/secondary-organic-aerosol-soas-research (October),

214. Xie, M.; Mladenov, N.; Williams, M. W.; Neff, J. C.; Wasswa, J.; Hannigan, M. P., Water soluble organic aerosols in the Colorado Rocky Mountains, USA: composition, sources and optical properties. *Sci. Rep.* **2016,** *6*, 39339.

215. Kirpes, R. M.; Bondy, A. L.; Bonanno, D.; Moffet, R. C.; Wang, B. B.; Laskin, A.; Ault, A. P.; Pratt, K. A., Secondary sulfate is internally mixed with sea spray aerosol and organic aerosol in the winter Arctic. *Atmos. Chem. Phys.* **2018,** *18*, (6), 3937-3949.

216. Eisenhauer, N., Lanoue, A., Strecker, T., Scheu, S., Steinauer, K., Thakur, M., Mommer, L., Root biomass and exudates link plant diversity with soil bacterial and fungal biomass. *Sci. Rep.* **2017,** *7*, (44641).

217. Gerhard, M., Dietmar, S., *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. 2nd ed. ed.; 2012.

218. EPA Understanding Global Warming Potentials. https://www.epa.gov/ghgemissions/understanding-global-warming-potentials (October 2018),

219. Knoblauch, C., Beer, C., Liebner, S., Grigoriev, M.N., Pfeiffer, E., Methane production as key to the greenhouse gas budget of thawing permafrost. *Nat. Clim. Change* **2018,** *8*, 309-312.

220. Lee, H.; Schuur, E. A. G.; Inglett, K. S.; Lavoie, M.; Chanton, J. P., The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. *Glob. Chang. Biol.* **2012,** *18*, (2), 515-527.

221. Ewing, S. O. D., J.A., Aiken, G.R., Butler, K., Butman, D., Windham-Myers, L., Kanevskiy, M.Z., Long-term anoxia and release of ancient, labile carbon upon thaw of Pleistocene permafrost. *Geophys. Res. Lett.* **2015,** *42*, (24).

222.   Rowland, J., Jones, C., Altmann, G., Bryan, R., Crosby, B., Hinzman, L., Kane, D., Lawrence, D., Mancino, A., March, P., McNamara, J., Romanovsky, V., Toniolog, H., Travis, B., Trochim, E., Wilson, C. Geernaert, G., Arctic landscapes in transition: responses to thawing permafrost. *Eos, Transactions American Geophysical Union* **2010,** *91*, (26).

223.   Hoffman, F. M.; Kumar, J.; Mills, R. T.; Hargrove, W. W., Representativeness-based sampling network design for the State of Alaska. *Landsc. Ecol.* **2013,** *28*, (8), 1567-1586.

224.   Your Weather Service U.S. Climate Data - Alaska. https://www.usclimatedata.com/climate/nome/alaska/united-states/usak0170

225.   Philippe, B., Mariette, J., Escudie, F., Djemiel, C., Klopp, C., jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **2014,** *15*, (293).

226.   Li, H.; Lee, L. S.; Jafvert, C. T.; Graveel, J. G., Effect of substitution on irreversible binding and transformation of aromatic amines with soils in aqueous systems. *Environ. Sci. Technol.* **2000,** *34*, (17), 3674-3680.

227.   Sakiroglu, H.; Ozturk, A. E.; Pepe, A. E.; Erat, M., Some kinetic properties of polyphenol oxidase obtained from dill (Anethum graveolens). *J. Enzyme Inhib. Med. Chem.* **2008,** *23*, (3), 380-385.

228.   Margalef, O.; Sardans, J.; Fernandez, F. M.; Molowny-Horas, R.; Janssens, I. A.; Ciais, P.; Goll, D.; Richter, A.; Obersteiner, M.; Asensio, D.; Penuelas, J., Global patterns of phosphatase activity in natural soils. *Sci. Rep.* **2017,** *7*, (1337).

229.   Banerjee, M. R.; Chapman, S. J., The significance of microbial biomass sulphur in soil. *Biol. Fertil. Soils* **1996,** *22*, (1-2), 116-125.

230.   Sardans, J.; Penuelas, J., Increasing drought decreases phosphorus availability in an evergreen Mediterranean forest. *Plant and Soil* **2004,** *267*, (1-2), 367-377.

231.   UNEP *Policy Implications of a Warming Permafrost*; United Nations Environment Program: 2012.

232.   Grosse, G., Harden, J., Tretsky, M., McGuire, A.D., Camill, P., Tarnocai, C., Frolking, S., Schuur, E. A. G., Joregnson, T., Marchenko, S., Romanovsky, V., Wickland, K.P., French, N., Waldrop, M., Bourgeau-Chavez, L., Striegl, R.G.,

Vulnerability of high latitude soil carbon in North America to disturbance. *JGR Biogeosciences* **2011,** *116*, (G00K06), 23.

233.   Tesar, C., Dubois, M., Shestakov, A. , Toward strategic, coherent, policy-relevant Arctic science. *Science* **2016,** *353*, (6306), 1368-1370.

234.   Rice, B., Richie, B, Schrooten, P., and Barnes, J. NPS Alaska Planning and Designs for the Future with Climate Change. https://www.nps.gov/articles/aps-v12-i2-c16.htm (September, 2018),

235.   NPS National Register of Historic Places Program. https://www.nps.gov/nr/feature/archeology/2012/climate_change.htm (September, 2018),

236.   Nixon, U. S. P., *Public Papers of the Presidents of the United States*. Government Printing Office: 1969-1974; p 1362.

237.   DOI, U. S. D. o. I. *Alaska National Interest Lands Conservation Act*; 1980.

238.   USARC United States Arctic Research Commission: About USARC. https://www.arctic.gov/about_usarc.html (September, 2018),

239.   DHS Homeland Security Presidential Directive 25: Arctic Region Policy. https://www.hsdl.org/?abstract&did=232474 (September, 2018),

240.   Obama, U. S. P. B. *National Strategy for the Arctic Region*; U.S. White House: 2013.

241.   United Nations Convention on the Law of the Sea. In Nations, U., Ed. 1982-1994.

242.   Ham, B., AAAS seeks to uphold science's role in policy-making. *Science* **2017,** *355*, (6332), 1383-1384.

243.   State, U. S. D. Defining the Limits of the U.S. Continental Shelf. https://2001-2009.state.gov/g/oes/continentalshelf/ (September),

244.   U.S.G.S. Bathymetric Surveys. https://www.usgs.gov/centers/oki-water/science/bathymetric-surveys?qt-science_center_objects=0#qt-science_center_objects (September),

245.   NOAA Arctic Charting: Mapping a New Frontier. https://www.noaa.gov/explainers/arctic-charting-mapping-new-frontier (September),

246.   NPS ANILCA and Alaska Wilderness. https://www.nps.gov/subjects/aknatureandscience/anilcawilderness.htm (September),

247.   Franzosa, E. A.; Hsu, T.; Sirota-Madi, A.; Shafquat, A.; Abu-Ali, G.; Morgan, X. C.; Huttenhower, C., Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* **2015,** *13*, (6), 360-372.

248.   Rajasundaram, D., Selbig, J., More effort - more results: Recent advances in integrative 'omics' data analysis. *Curr. Opin. Plant Biol.* **2016,** *30*, 57-61.

249.   Li, Z.; Yao, Q. M.; Dearth, S. P.; Entler, M. R.; Castro Gonzalez, H. F.; Uehling, J. K.; Vilgalys, R. J.; Hurst, G. B.; Campagna, S. R.; Labbe, J. L.; Pan, C. L., Integrated proteomics and metabolomics suggests symbiotic metabolism and multimodal regulation in a fungal-endobacterial system. *Environ. Microbiol.* **2017,** *19*, (3), 1041-1053.

250.   Graham, E. B.; Crump, A. R.; Kennedy, D. W.; Arntzen, E.; Fansler, S.; Purvine, S. O.; Nicora, C. D.; Nelson, W.; Tfaily, M. M.; Stege, J. C., Multi 'omics comparison reveals metabolome biochemistry, not microbiome composition or gene expression, corresponds to elevated biogeochemical function in the hyporheic zone. *Sci. Total Environ.* **2018,** *642*, 742-753.

251.   Gandhi, A., Shah, N., Integrating omics to unravel the stress-response mechanisms in probiotic bacteria: Approaches, challenges, and prospects. *Crit. Rev. Food Sci. Nutr.* **2017,** *57*, (16), 3464-3471.

252.   Raja, K.; Patrick, M.; Gao, Y. L.; Madu, D.; Yang, Y. Y.; Tsoi, L. C., A Review of Recent Advancement in Integrating Omics Data with Literature Mining towards Biomedical Discoveries. *International Journal of Genomics* **2017**.

253.   Merelli, I.; Perez-Sanchez, H.; Gesing, S.; D'Agostino, D., Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. *Biomed Research International* **2014**.

254.   Zeng, I. S. L., Lumley, T., Review of statistical learning methods in integrated omics studies. *Bioinform. Biol. Insights* **2018,** *12*.

255.   Huang, S.; Chaudhary, K.; Garmire, L. X., More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Frontiers in Genetics* **2017,** *8*, (84).

256.   Hultman, J., Waldrop, M., Mackelprang, R., David, M., McFarland, J., Blazewicz, S., Harden, J., Turetsky, M., McGuire, A. D., Shah, M., VerBerkmoes, N., Lee, L.,

Mavrommatis, K., Jansson, J., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **2015,** *521*, 208-212.

257.   Mackelprang, R.; Saleska, S. R.; Jacobsen, C. S.; Jansson, J. K.; Tas, N., Permafrost Meta-Omics and Climate Change. *Annual Review of Earth and Planetary Sciences, Vol 44* **2016,** *44*, 439-462.

258.   Lutz, S., Anesio, A., Field, K., Benning, L., Integrated 'Omics', Targeted Metabolite and Single-Cell Analyses of Arctic Snow Algae Functionality and Adaptability. *Front. Microbiol.* **2015,** *6*, 1323.

259.   Bell, T., Yergeau, E., Maynard, C., Juck, D., Whyte, L., Greer, C., Predictable bacterial composition and hydrocarbon degradation in ARctic soils following diesel and nutrient disturbance. *ISME* **2013,** *7*, 1200-1210.

260.   Gonzalez-Martinez, A., Sihvonen, M., Munoz-Palazon, B., Microbial ecology of full-scale wastewater treatment systems in the Polar Arctic Circle: Archaea, Bacteria, and Fungi. *Sci. Rep.* **2018,** *8*, (1), 2208.

261.   Byrnes, M. R., Baker, J. L., Li, F. *Quantifying Potential Measurement Errors and Uncertainties Associated with Bathymetric Change Analysis*; 2002.

262.   M Bollmann, T. B., F Colijn, R Ebinghaus, R Froese, K Guessow, S Khalilian, S Krastel, A Koertzinger, M Lagenbuch, M Latif, B Matthiessen, F Melzner, A Oschlies, S Petersen, A Proelss, M Quaas, J Reichenbach, T Requate, T Reusch, P Rosenstiel, JO Schmidt, K Schrottke, H Sichelschmidt, U Siebert, R Soltwedel, U Sommer, K Stattegger, H Sterr, R Sturm, T Truede, A Vafeidis, C Van Bernem, J Van Beusekom, R Voss, M Visbeck, M Wahl, K Wallmann, F Weinberger *Living with the oceans: A report on the state of the world's oceans*; maribus, 2010.

263.   McGlade, C. E., A review of the uncertatinties in estimates of global oil resources. *Energy* **2012,** *47*, (1), 262-270.

264.   Joppa, L., Pfaff, A., High and far: Biases in the location of protected areas. *PLoS One* **2009,** *4*, (12), e8273.

265.   Stuart H. M. Butchart , J. P. W. S., Mike I. Evans, Suhel Quader, Salvatore Aricò, Julius Arinaitwe, Mark Balman, Leon A. Bennun, Bastian Bertzky, Charles Besançon, Timothy M. Boucher, Thomas M. Brooks, Ian J. Burfield, Neil D. Burgess, Simba Chan,

Rob P. Clay, Mike J. Crosby, Nicholas C. Davidson, Naamal De Silva, Christian Devenish, Guy C. L. Dutson, David F. Día z Fernández, Lincoln D. C. Fishpool, Claire Fitzgerald, Matt Foster, Melanie F. Heath, Marc Hockings, Michael Hoffmann, David Knox, Frank W. Larsen, John F. Lamoreux, Colby Loucks, Ian May, James Millett, Dominic Molloy, Paul Morling, Mike Parr, Taylor H. Ricketts, Nathalie Seddon, Benjamin Skolnik, Simon N. Stuart, Amy Upgren, Stephen Woodley, Protecting Immportant Sites for Biodiversity Contributes to Meeting Global Conservation Targets. *PLoS One* **2012,** *7*, (3), e32529.

266.   Arriagada, R., Echeverria, C., Moya, D., Creating protected areas on public lands: Is there room for additional conservation? *PLoS One* **2016,** *11*, (2).

267.   Gillingham, P., Bradbury, R., Roy, D., Anderson, B., Baxter, J., Bourn, N., Crick, H., Findon, R., Fox, R., Franco, A., Hill, J., Hodgson, J., Holt, A., Morecroft, M., O'Hanlon, N., Oliver, T., Pearce-Higgins, J., Procter, D., Thomas, J., Walker, K., Walmsley, C., Wilson, R., Thomas, C., The effectiveness of protected areas in the conservation of species with changing geographical ranges. *Biol. J. Linn. Soc.* **2015,** *115*, (3), 707-717.

268.   da Silva, M. X., Paviolo, A., Tambosi, L.R., Pardini, R., Effectiveness of protected areas for biodiversity conservation: Mammal occupancy patterns in the Iguacu National Park, Brazil. *J. Nat. Conserv.* **2018,** *41*, 51-62.

269.   Dihazi, H.; Asif, A. R.; Beißbarth, T.; Bohrer, R.; Feussner, K.; Feussner, I.; Jahn, O.; Lenz, C.; Majcherczyk, A.; Schmidt, B.; Schmitt, K.; Urlaub, H.; Valerius, O., Integrative omics - from data to biology. *Expert Review of Proteomics* **2018,** *15*, (6), 463-466.

270.   Joyce, A. R., Palsson, B.O., The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Molec. Cell Biol.* **2006,** *7*, (3), 198-210.

271.   Karczewski, K. J.; Snyder, M. P., Integrative omics for health and disease. *Nat. Rev. Genetics* **2018,** *19*, 299.

272.   Kohlbacher, O.; Vitek, O., Challenges in Large-Scale Computational Mass Spectrometry and Multiomics. *J. Proteome Res.* **2016,** *15*, (3), 681-682.

273.   Zhang, W., Li, F., Nie, L., Integrating multiple 'omics' analysis for microbial biology: applications and methodologies. *Microbiology* **2010,** *156*, 287-301.

274.   Zhou, Z.; Gu, J.; Du, Y.-L.; Li, Y.-Q.; Wang, Y., The -omics Era- Toward a Systems-Level Understanding of Streptomyces. *Current Genomics* **2011,** *12*, (6), 404-416.

275.   Tonn, B., Healy, K.C., Gibson, A., Ashish, A., Cody, P., Beres, D., Lulla, S., Mazur, J., Ritter, A.J., Power from Perspective: Potential Future United States Energy Portfolios. *Energy Policy* **2009,** *37*, 1432-1443.

276.   Kerr, R., A Final Push to Divvy Up the Sea by All the Rules. *Science* **2009,** *324*, (5934), 1500-1501.

277.   Hirono, M. K., Hirono, Murkowski Introduce Bipartisan Resolution Urging Approval of Law of the Sea Treaty. In *July Marks 24 Years Since U.S. Signed Treaty But Senate Has Failed to Vote on Ratification*, Hirono Website, 2018.

278.   Johnson, H., Why the US Needs to Ratify UNCLOS. *The Diplomat* April 18, 2016, 2016.

279.   Houck, J., The Opportunity Costs of Ignoring the Law of Sea Convention in the Arctic. In *Arctic Security Initiative*, Institution, H., Ed. Penn State Law eLibrary: 2013.

280.   Sykes, A., Posner, E. *Economic Foundations of the Law of the Sea*; University of Chicago Law School: Chicago Unbound, 2009.

281.   Johnson, J., Letter from Admiral Jay Johnson, U.S. Navy, Chief of Naval Operations, to Senator Jesse Helms, Chairperson, U.S. Senate Committee on Foreign Relations. In Senate, U. S., Ed. Naval Justice School: Naval Law Review, 2000; Vol. LXI.

282.   U.S.J.A.G. The Convention on the Law of the Sea. http://www.jag.navy.mil/organization/code_10_law_of_the_sea.htm (September),

283.   Hearing on the Implementation of the Alaska National Interest Lands Conservation Act of 1980. In *Senate Energy and Natural Resources Committee*, First ed.; Committee on Energy and Natural Resources: https://www.gpo.gov/fdsys/pkg/CHRG-114shrg98964/pdf/CHRG-114shrg98964.pdf, 2015.

284.	ADNR State ANILCA Coordination: Mission and Objectives. http://dnr.alaska.gov/commis/opmp/anilca/ (September),

285.	Begich, S. M., Arctic Research, Monitoring, and Observing Act of 2012. In Senate, U. S., Ed. Committee on Commerce, Science, and Transportation: congress.gov, 2012.

286.	Gilens, M., Page, B., Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. *Perspectives on Politics* **2014,** *12*, (3), 564-581.

287.	Bird, K., Charpentier, R., Gautier, D., Houseknecht, D., Klett, T., Pitman, J., Moore, T., Schenk, C., Tennyson, M. Wandrey, C. *Circum-Arctic Resource Appraisal: Estimates of Undiscovered Oil and gas North of the Arctic Circle*; USGS: 2008.

288.	EIA *Arctic oil and natural gas resources*; 2012.

289.	Alaska, T. f. Celebrating 35 Years of ANILCA: Monumental Conservation Act Shaped the Alaska We Know Today. http://www.trustees.org/celebrating-35-years-of-anilca/ (September),

290.	Ristroph, E., Hussain, A., Wilderness: Good for Alaska. *Washington Journal of Environmental Law & Policy* **2015**.

291.	NPS *National Park Visitor Spending Effects: Economic Contributions to Local Communities, States, and the Nation*; 2012.

292.	Vaughn, D. G., Comiso, J.C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffan, K., Zhang, T. *Observations: Cryosphere. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK and NY, USA, 2013.

293.	Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L.-F.; Dührkop, K.; Bandeira, N.; Dorrestein, P. C.; Böcker, S., Significance estimation for large scale metabolomics annotations by spectral matching. *Nature Comm.* **2017,** *8*, (1), 1494.

APPENDICES

Appendix A: Bioinformatic workflow

Peak Detection Methods:

CA_Sample_2_HILIC_pos.raw scan#1 @0.01 MS1 c +, base peak: 471.7955 m/z (2.0E5)
Scan definition: FTMS + c NSI Full ms [50.00-1000.00]
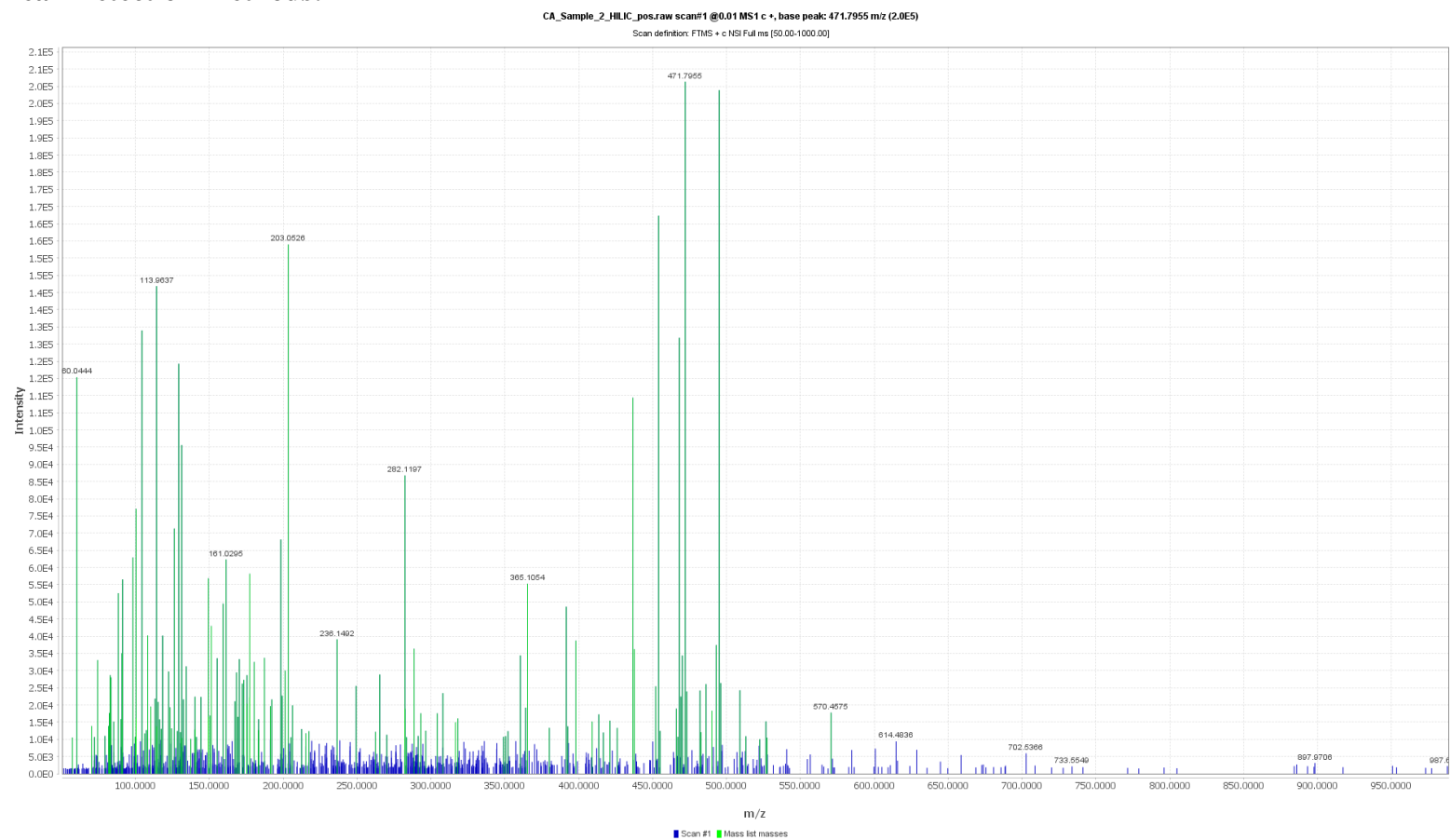
Figure 53: *Mass Detection:* Generates a list of masses (ions) for each scan in the raw data file using the centroid mass detector algorithm which assigns peaks above a given noise level (shown in blue)
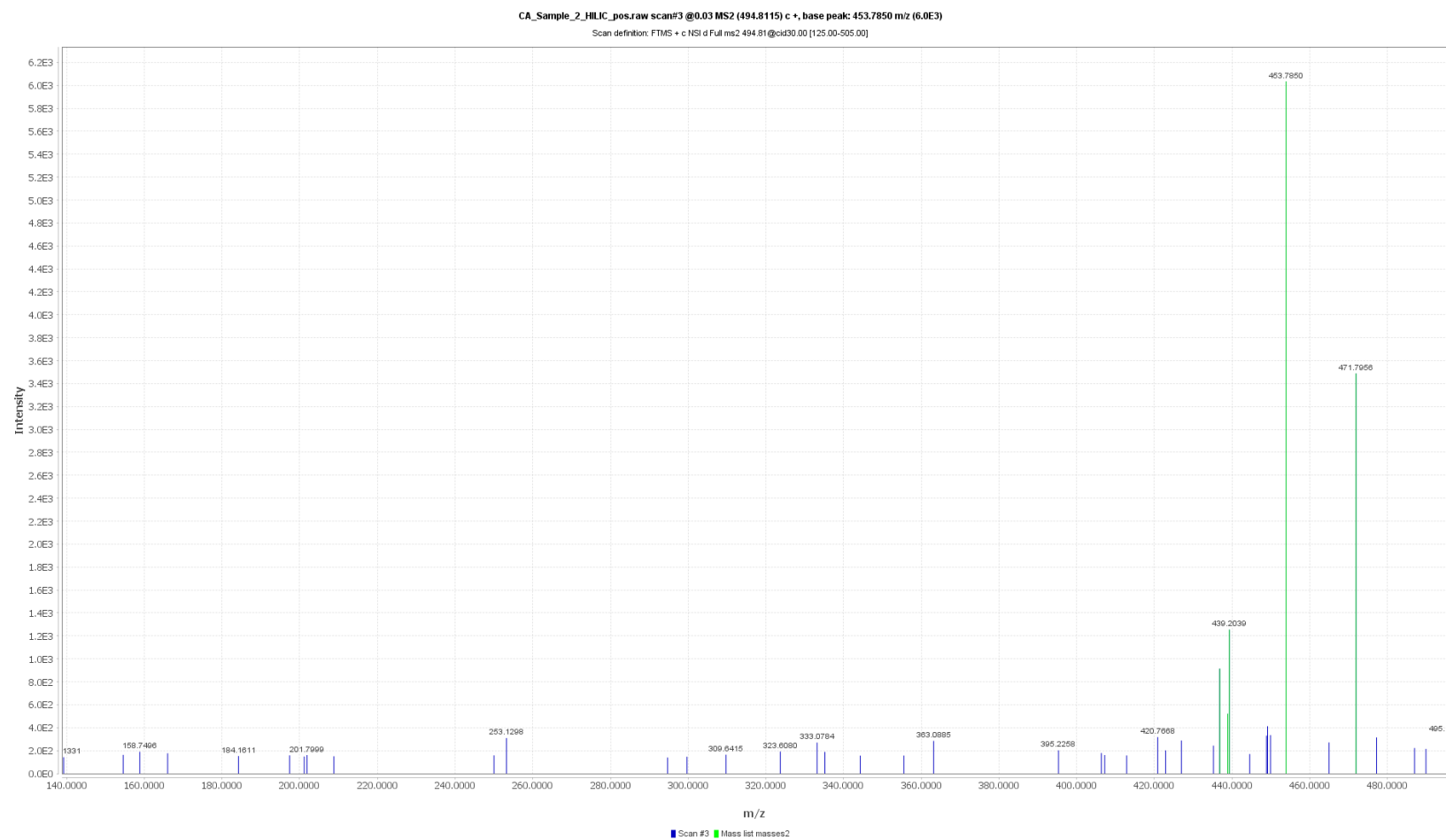
Figure 54: Example of MS^2 spectrum with green indicating peaks selected for analysis and blue indicating peaks that were discarded (excluded from downstream analyses)

ID ∧	Average		Identity	Comment	Peak shape	CA_Sample_2_HILIC_pos.raw		
	m/z	RT				Status	Height	Area
2	471.7955	4.23				●	9.3E5	9.3E5
3	494.8115	4.23				●	9.9E5	9.9E5
4	453.7846	4.23				●	7.9E5	7.9E5
5	203.0526	6.91				●	8.1E5	8.1E5
6	113.9600	4.69				●	2.2E5	2.2E5
8	104.1069	5.66				●	4.4E6	4.4E6
9	467.8007	5.82				●	2.2E5	2.2E5
10	128.9500	3.91				●	2.3E5	2.3E5
11	60.0400	3.41				●	1.9E5	1.9E5
12	436.3418	1.69				●	1.6E5	1.6E5
14	130.9700	4.30				●	1.3E5	1.3E5
15	100.0800	2.52				●	4.8E5	4.8E5
16	97.9900	0.74				●	6.8E4	6.8E4
17	90.9500	4.23				●	9.0E4	9.0E4
18	365.1052	10.39				●	1.9E5	1.9E5
20	149.0233	0.74				●	8.9E4	8.9E4
21	125.9862	3.66				●	1.1E5	1.1E5
22	282.1196	5.19				●	4.6E6	4.6E6
23	158.9600	24.65				●	1.6E5	1.6E5
24	198.0972	7.88				●	6.5E5	6.5E5

Figure 55: *MS/MS Peak List Builder*: Searches raw data for MS^2 scans, then makes a list of parent scans (MS^1) that have fragmentation data and builds a chromatogram at the retention time with a corresponding peak height and area
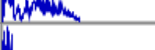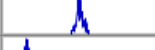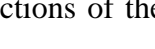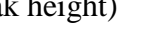
ID ^	Average		Identity	Comment	Peak shape	CA_Sample_2_HILIC_pos.raw		
	m/z	RT				Status	Height	Area
2	471.7953	4.23				🟢	9.3E5	3.0E8
3	494.8113	4.23				🟢	9.9E5	4.0E8
4	453.7846	4.23				🟢	7.9E5	2.6E8
5	203.0526	6.91				🟢	8.1E5	4.2E8
6	113.9636	4.69				🟢	2.2E5	6.7E7
8	104.1069	5.66				🟢	4.4E6	5.9E8
9	467.8008	5.82				🟢	2.2E5	2.5E8
10	128.9507	3.91				🟢	2.3E5	6.3E7
11	60.0444	3.41				🟢	1.9E5	9.8E7
12	436.3418	1.69				🟢	1.6E5	2.1E7
14	130.9664	4.30				🟢	1.3E5	4.1E7
15	100.0756	2.52				🟢	4.8E5	1.7E7
16	97.9913	0.74				🟢	6.8E4	2.0E7
17	90.9476	4.23				🟢	9.0E4	1.5E7
18	365.1051	10.39				🟢	1.9E5	3.5E7
20	149.0233	0.74				🟢	8.9E4	3.7E7
21	125.9862	3.66				🟢	1.1E5	1.3E7
22	282.1196	5.19				🟢	4.6E6	4.2E8
23	158.9640	24.65				🟢	1.6E5	1.9E7
24	198.0972	7.88				🟢	6.5E5	1.6E8

Figure 56: *Peak Extender:* Extends the chromatographic peak in both directions of the apex retention using a scan-by-scan search within a given m/z tolerance and above a given minimal intensity (peak height)

Peak List Methods:

Isotopic Peaks Grouper: Searches for isotopes within peak list (1.0033 Da away within user-defined *m/z* tolerance). Most intense isotope is kept, and others are removed from peak list (no figure).
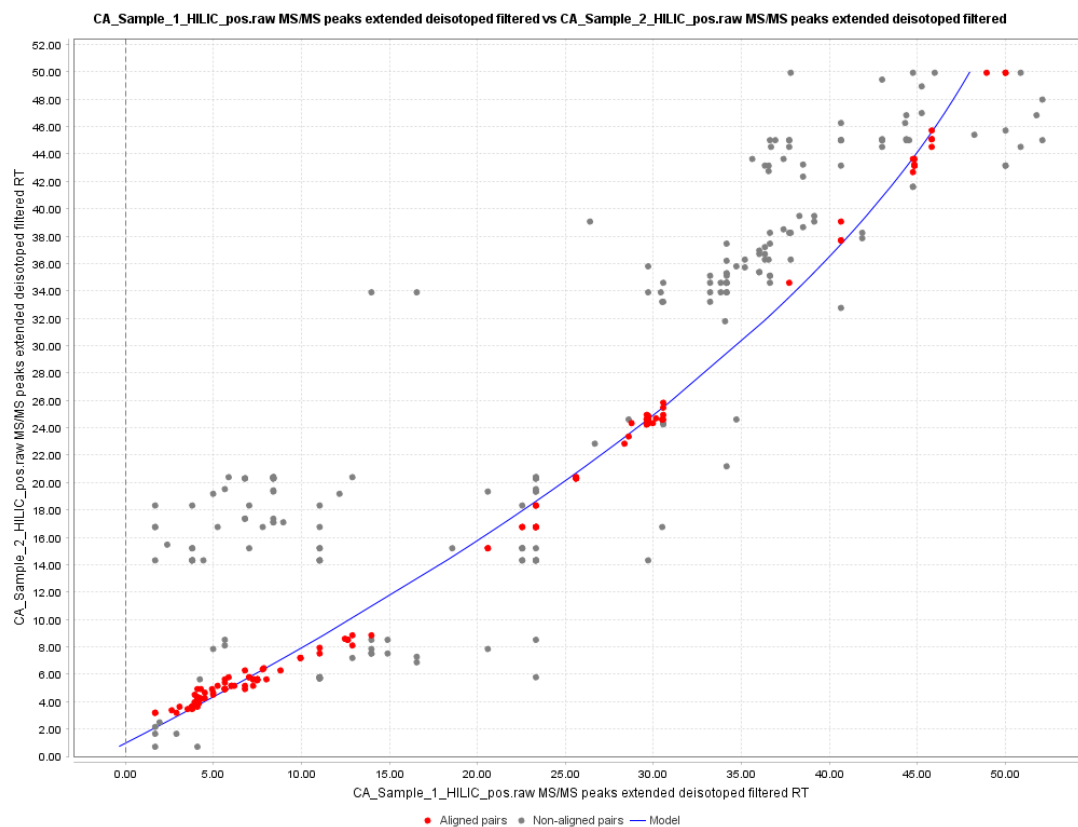
Figure 57: *RANSAC Aligner*: Aligns chromatograms in peak lists across samples, correcting for any linear or non-linear RT deviations (within a designated threshold) by creating a nonlinear regression model of the features and their deviations.

ID	Average		Peak shape	CA_Blank_HILIC_pos_05122017.raw			CA_H2O_Control_2_HILIC_pos.raw			CA_H2O_Control_3_HILIC_pos.raw		
	m/z	RT		Status	Height	Area	Status	Height	Area	Status	Height	Area
1	111.0916	12.43		🟢	9.3E5	5.4E8	🟢	8.4E4	1.3E7	🟢	1.6E6	7.1E8
2	494.8118	10.70		🟢	4.2E5	5.3E8	🟢	5.7E4	1.7E6	🟢	1.1E6	5.5E8
3	471.7960	26.04		🟢	3.0E5	3.8E8	🟢	9.7E4	5.3E7	🟢	2.4E5	2.7E6
4	88.0756	14.10		🟢	3.4E5	3.7E8	🟡	1.0E5	6.3E7	🟢	1.5E6	2.7E8
5	453.7850	25.64		🟢	3.0E5	3.7E8	🟢	1.1E5	5.4E7	🟡	7.0E5	8.0E7
6	74.0963	12.14		🟢	6.1E5	3.6E8	🟡	3.9E4	8.6E6	🟢	7.2E5	1.4E8
7	467.8012	24.36		🟢	2.7E5	3.4E8	🟢	2.2E5	1.9E8	🟢	5.1E5	5.5E8
8	104.1069	8.59		🟢	4.0E5	3.2E8	🟢	3.5E4	2.8E6	🟢	2.7E6	4.2E8
9	113.9636	36.13		🟢	3.1E5	3.2E8	🟢	1.3E6	1.9E9	🟢	2.1E5	2.0E8
10	128.9508	35.43		🟢	2.8E5	2.9E8	🟢	7.9E5	1.1E9	🟢	2.0E5	2.0E8
11	139.9879	35.69		🟢	3.7E5	2.9E8	🟢	2.1E5	1.9E8	🟢	4.2E5	3.3E8
12	60.0443	16.09		🟢	1.6E5	2.3E8	🟢	1.8E5	2.7E8	🟢	1.5E5	1.9E8
13	288.2516	10.68		🟢	2.4E5	2.2E8	🟢	7.8E5	4.5E8	🟢	3.7E5	2.8E8
14	436.3425	34.41		🟢	3.9E5	2.0E8	🟢	6.4E5	4.5E8	🟢	3.5E5	2.3E8
15	139.1230	11.17		🟢	4.4E5	1.9E8	🟡	1.9E4	1.4E6	🟢	5.2E5	2.4E8
16	282.1198	6.88		🟢	1.5E5	1.9E8	🟢	4.8E6	5.2E8	🟢	9.3E5	3.2E8
17	130.9664	35.94		🟢	1.8E5	1.8E8	🟢	5.3E5	7.7E8	🟢	1.4E5	1.2E8

Figure 58: *Gap Filling*: Fills in gaps in aligned peak list by looking for entries that fell outside the RT tolerance but fell within a tighter *m/z* tolerance. Reduces false negatives due to RT variation after alignment.

ID ^	Average		Identity	Comment	Peak shape	CA_Blank_HILIC_pos_05122017.raw		
	m/z	RT				Status	Height	Area
286	344.3125	11.80				●	4.2E4	5.5E5
287	318.7919	21.37				●	2.0E4	5.3E5
288	142.9665	27.92	[M+FA-H] 44.9982 m/z adduct of 97.9686 m/z			●	1.8E4	5.0E5
289	492.8164	6.22				●	4.2E4	4.8E5
290	160.9856	25.22				●	2.2E4	4.8E5
291	344.2264	12.47				●	2.3E4	4.7E5
292	204.9661	25.75				●	1.6E4	4.7E5
293	436.3991	16.10				●	3.5E4	4.6E5
294	100.5098	31.04				●	1.7E4	4.6E5
295	397.7954	6.95	Fragment of 453.7847 m/z			●	3.3E4	4.5E5
296	350.2175	23.54				●	3.3E4	4.4E5
297	96.9606	30.06				●	2.0E4	4.4E5
298	156.9901	32.11				●	1.6E4	4.2E5
299	307.9142	32.25				●	1.8E4	4.2E5
300	205.9873	24.85				●	1.7E4	4.1E5
301	244.9467	32.17				●	2.5E4	4.1E5
302	448.6989	35.52				●	1.8E4	4.0E5
303	331.8442	35.53				●	2.0E4	3.9E5
304	90.5257	20.83	Fragment of 351.9400 m/z			●	2.1E4	3.7E5
305	310.8471	19.31				●	1.9E4	3.7E5
306	223.9885	28.66				●	2.6E4	3.5E5
307	644.4931	22.13				●	2.4E4	3.5E5
308	262.1649	6.82	Complex of 113.9636 and 171.1860 m/z			●	1.8E4	3.5E5

Figure 59: *Identification of fragments, adducts, and complexes*: Searches peak lists using MS^2 data, RT and *m/z* thresholds

Modules annotate features that appear to be 1) fragments at the same retention time using MS/MS scan data within an *m/z* tolerance 2) adducts formed by the interaction of two ions (i.e. salt ions, water) using a common built-in list of adducts, and 3) pairs of ions that appear at the same retention time and form an ion complex in the spectrum containing both smaller ions as components.

ID	Average m/z	Average RT	Identity	Comment	Peak shape	CA_Blank_HILIC_pos_05122017.raw Status	Height	Area	CA_H2O_Control_2_HILIC_pos.raw Status	Height	Area
1	111.0916	12.43	1,3,5-TRIMETHYLPYRAZOLE			●	9.3E5	5.4E8	●	8.4E4	1.3E7
2	494.8118	10.70	MolPort-004-936-920			●	4.2E5	5.3E8	●	5.7E4	1.7E6
3	471.7960	26.04	SCHEMBL14509031			●	3.0E5	3.8E8	●	9.7E4	5.3E7
4	88.0756	14.10	4-Aminobutyraldehyde			●	3.4E5	3.7E8	●	1.0E5	6.3E7
5	453.7850	25.64	AC1NDDLO			●	3.0E5	3.7E8	●	1.1E5	5.4E7
6	74.0963	12.14	2-Methylpropanamine			●	6.1E5	3.6E8	●	3.9E4	8.6E6
7	467.8012	24.36	1-(4-chlorophenyl)-2-(3,5-dibromopyridin-1-ium-...			●	2.7E5	3.4E8	●	2.2E5	1.9E8
8	104.1069	8.59	2-isopropylaminoethanol			●	4.0E5	3.2E8	●	3.5E4	2.8E6
9	113.9636	36.13	AC1O4CTB			●	3.1E5	3.2E8	●	1.3E6	1.9E9
10	128.9508	35.43	Dichloroacetate			●	2.8E5	2.9E8	●	7.9E5	1.1E9
11	139.9879	35.69	acetyl hydrogen phosphate			●	3.7E5	2.9E8	●	2.1E5	1.9E8
12	60.0443	16.09	Acetamide			●	1.6E5	2.3E8	●	1.8E5	2.7E8
13	288.2516	10.68	Tricetanidin			●	2.4E5	2.2E8	●	7.8E5	4.5E8
14	436.3425	34.41	AC1L1TKG			●	3.9E5	2.0E8	●	6.4E5	4.5E8
15	139.1230	11.17	3088-41-3			●	4.4E5	1.9E8	●	1.9E4	1.4E6
16	282.1198	6.88	1-Methyladenosine			●	1.5E5	1.9E8	●	4.8E6	5.2E8
17	130.9664	35.94	2,2-Dichloro-1,1-ethanediol			●	1.8E5	1.8E8	●	5.3E5	7.7E8
18	100.0756	41.27	N-Methyl-2-pyrrolidinone			●	5.4E5	1.7E8	●	5.2E5	5.6E6

Figure 60: *Annotation with online databases:* MZmine annotation module searches a selected database (here: KEGG, PubChem, HMDB, LipidMaps, and Plant Cyc) for [M+H]$^+$ or [M-H]$^-$ ions within a 0.001 *m/z* or 5 ppm mass tolerance. The module returns the top 10 matches and exports them to an .csv matrix.

Appendix B: Fragmentation spectra
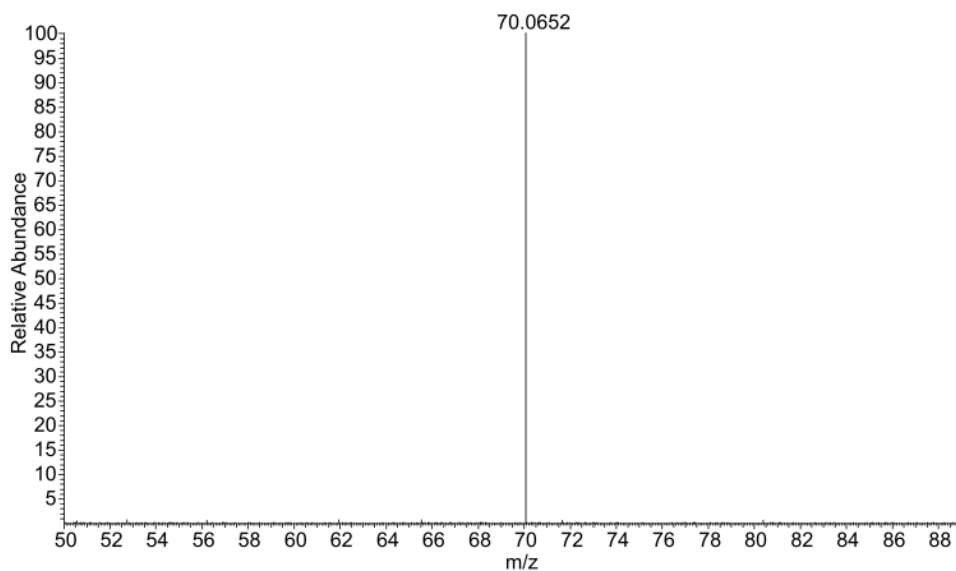
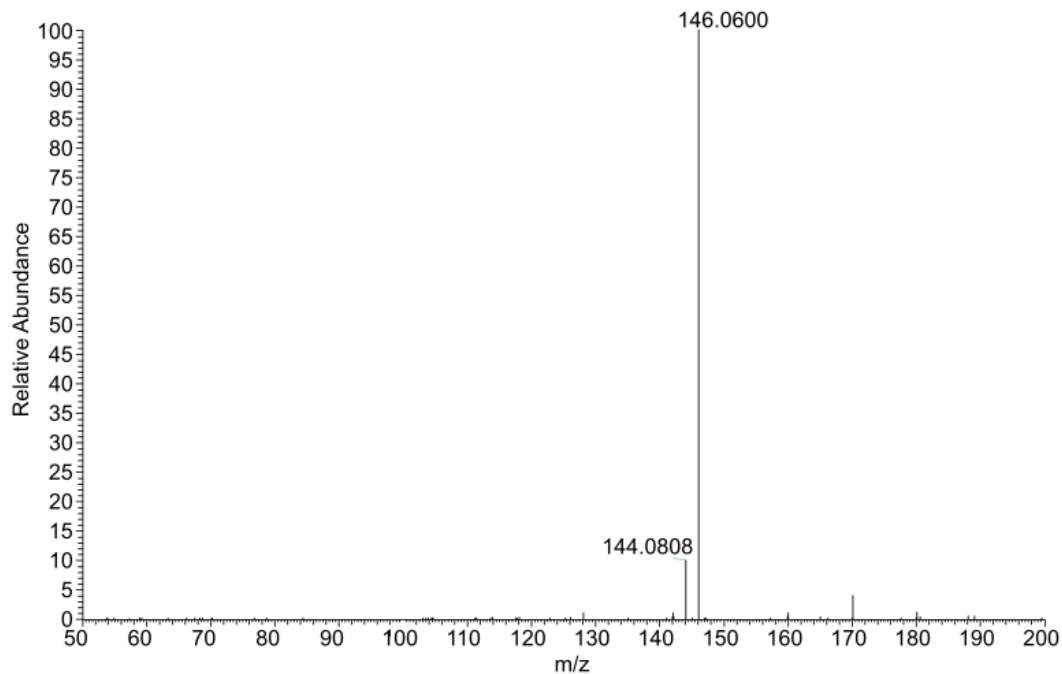**Figure 61:** Experimental MS^2 spectrum for feature detected by HILIC (+) at 116.0705 *m/z*
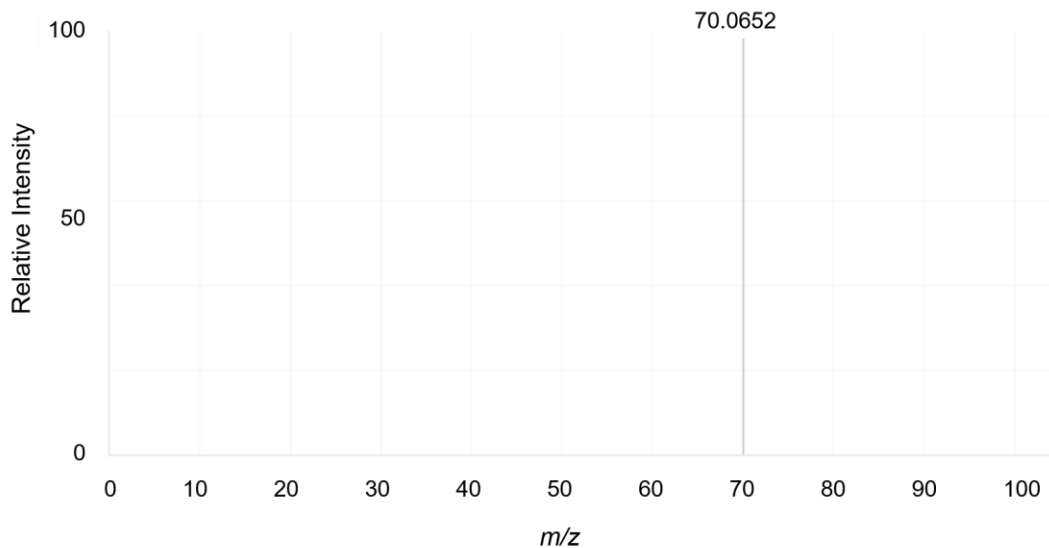
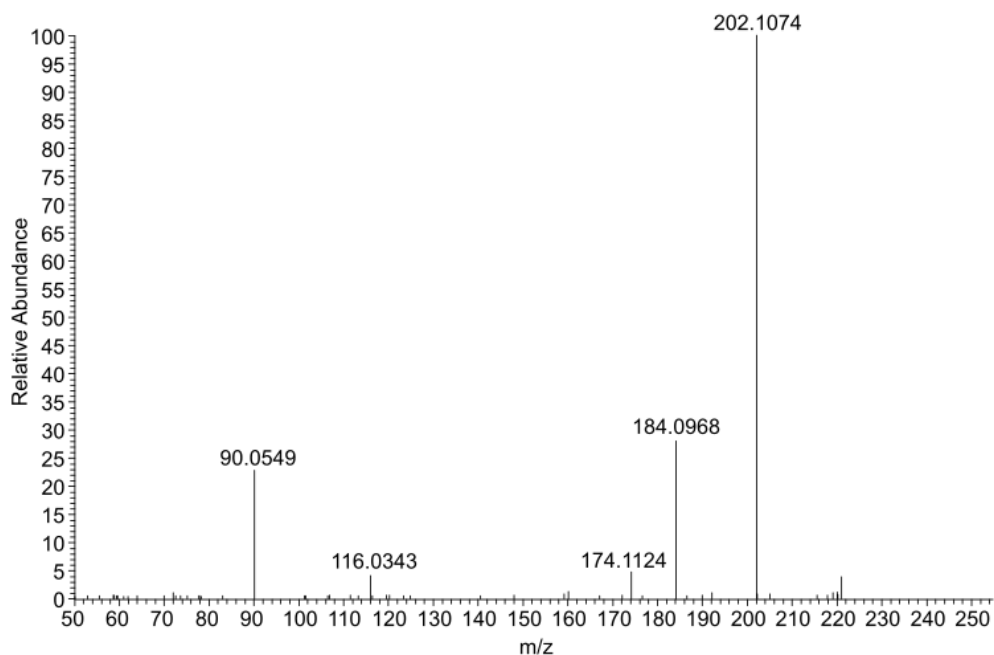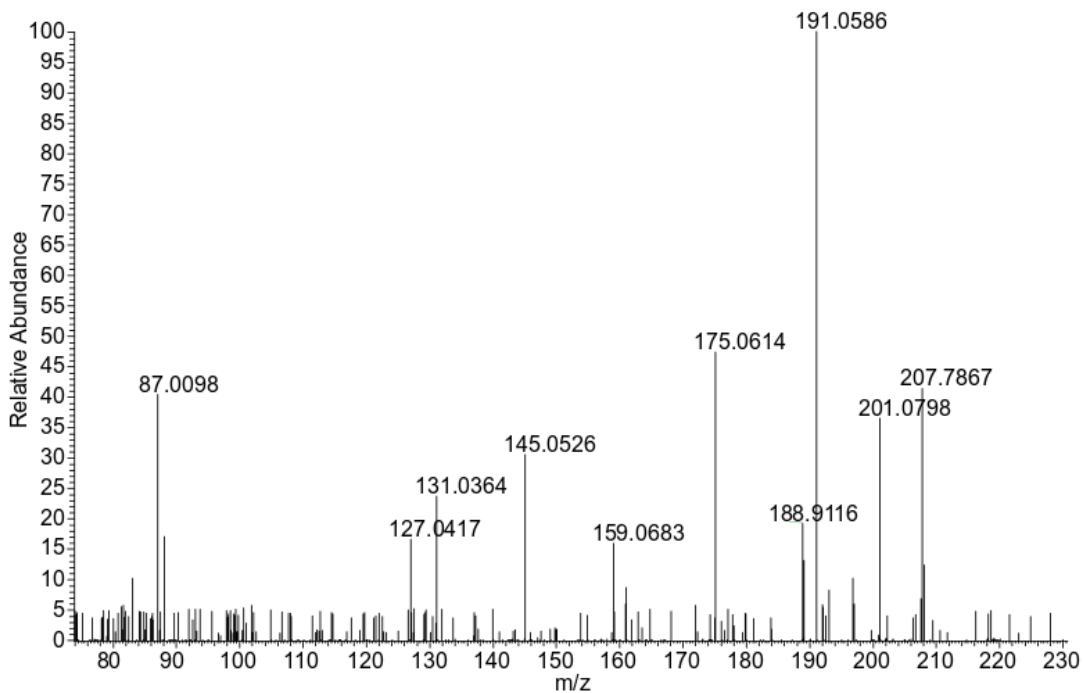**Figure 62:** Corresponding database MS^2 spectrum for matched standard, proline. Note: Databases where matched compound information were obtained from are listed in Table 5.
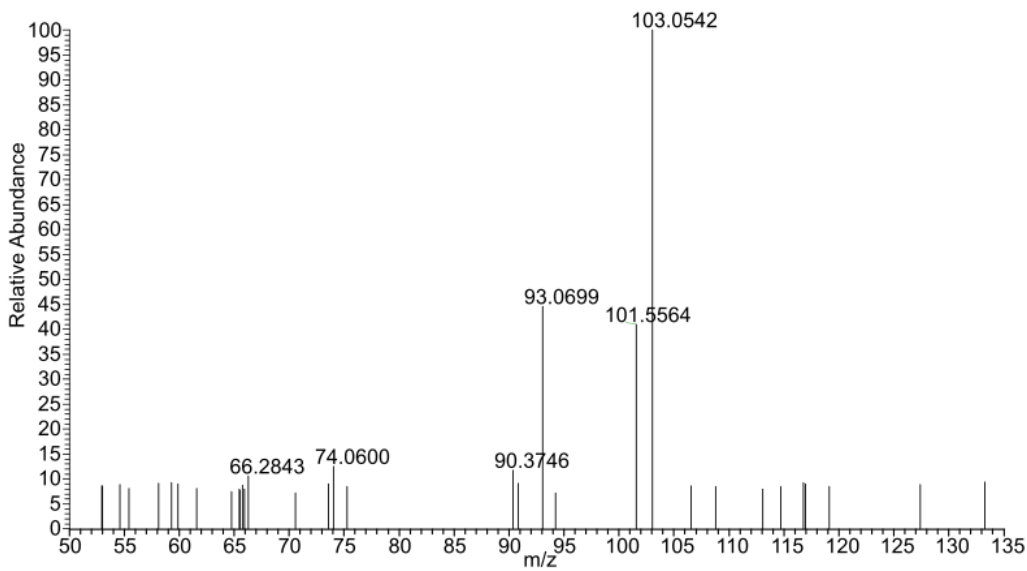
Figure 63: Experimental MS^2 spectrum for feature detected by HILIC (+) at 120.0807 *m/z*
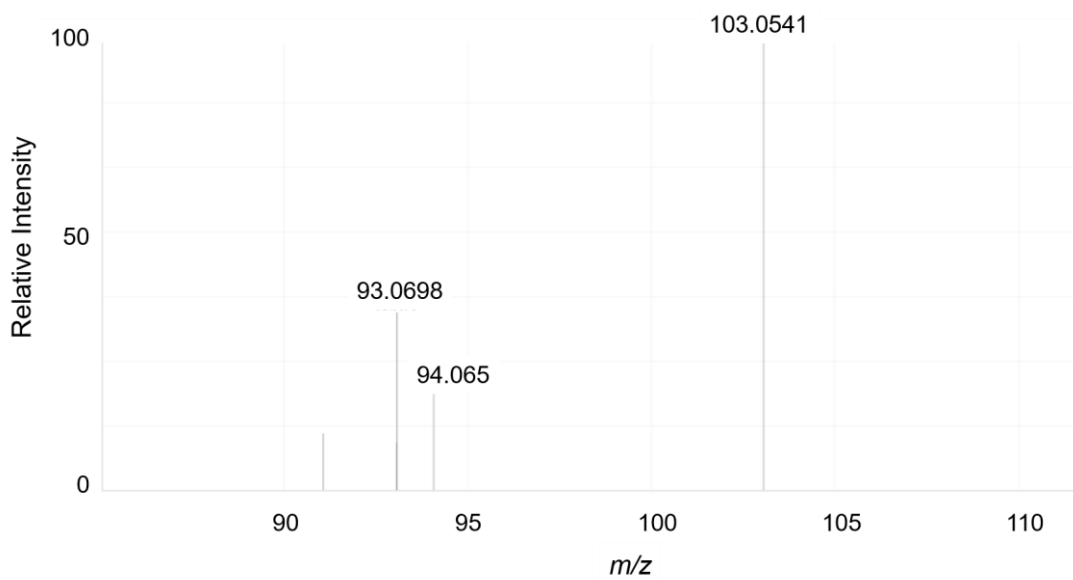
Figure 64: Corresponding database MS^2 spectrum for matched standard, indoline

232
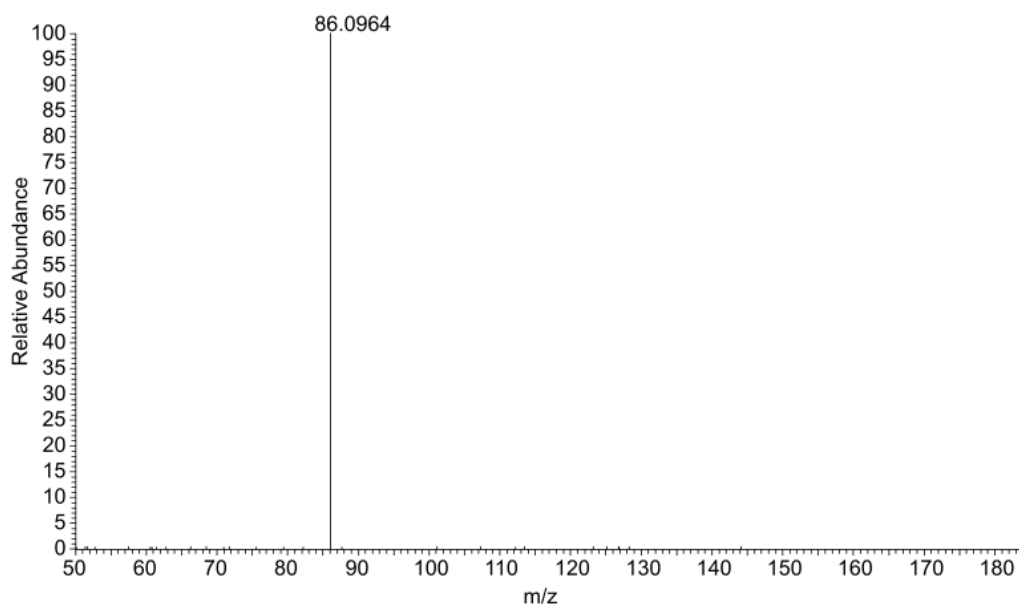
Figure 65: Experimental MS^2 spectrum for feature detected by HILIC (+) at 132.1018 *m/z*
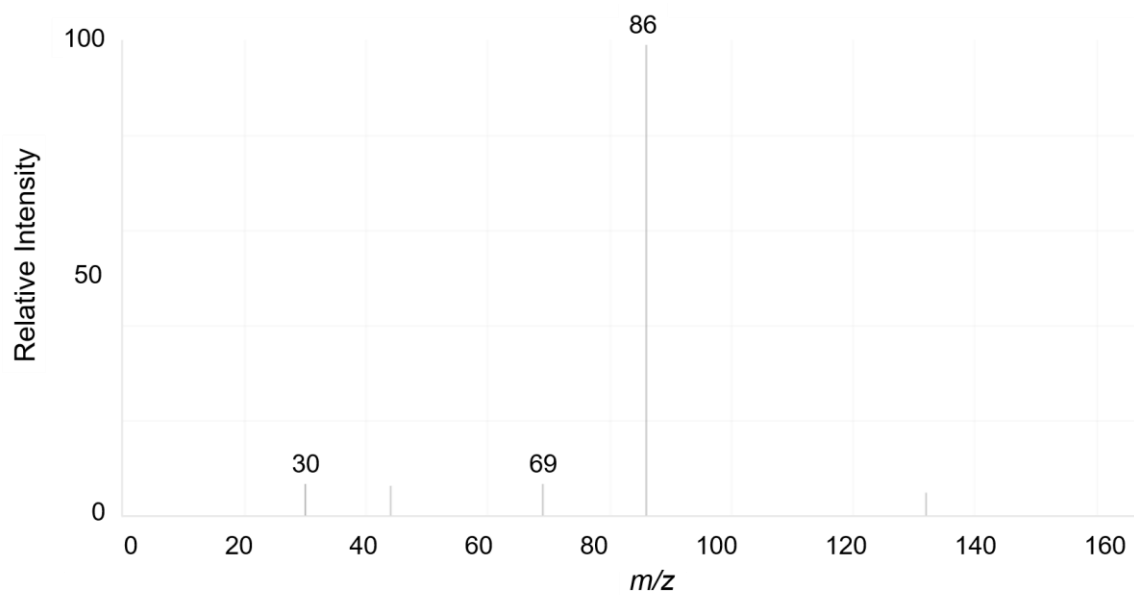
Figure 66: Corresponding database MS^2 spectrum for matched standard, alloisoleucine
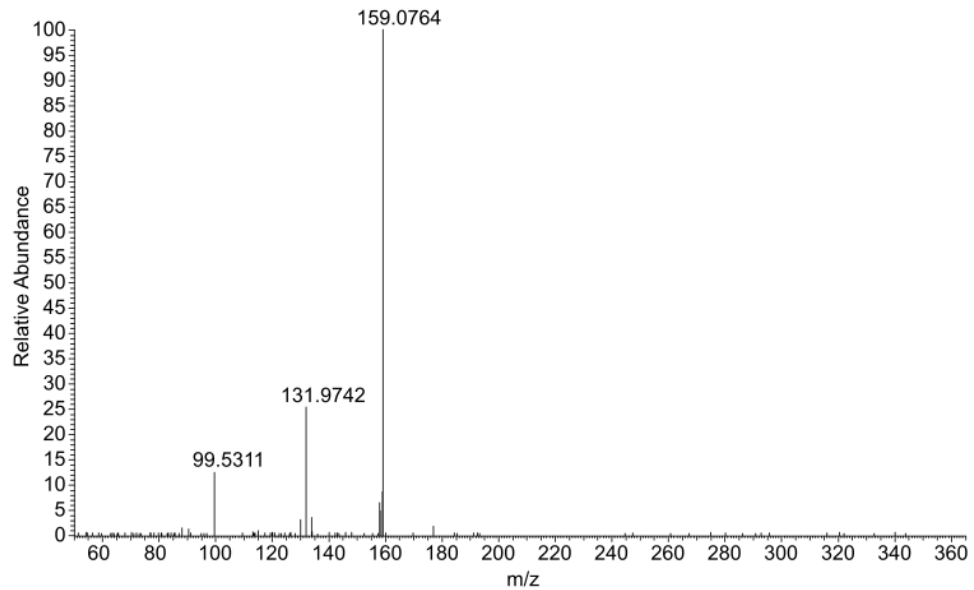
Figure 67: Experimental MS^2 spectrum for feature detected by HILIC (+) at 176.1028 *m/z*
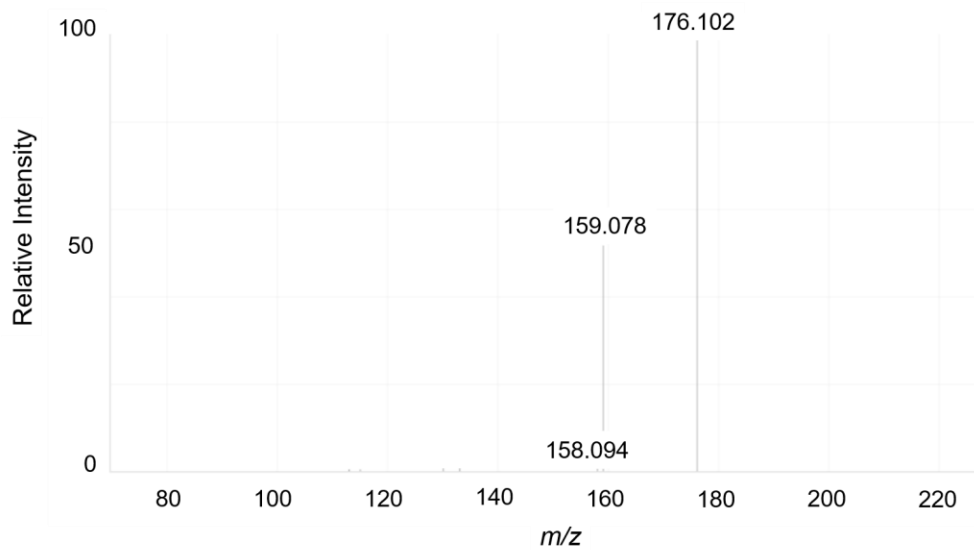
Figure 68: Corresponding database MS^2 spectrum for matched standard at lower CID energy (30 CID), citrulline
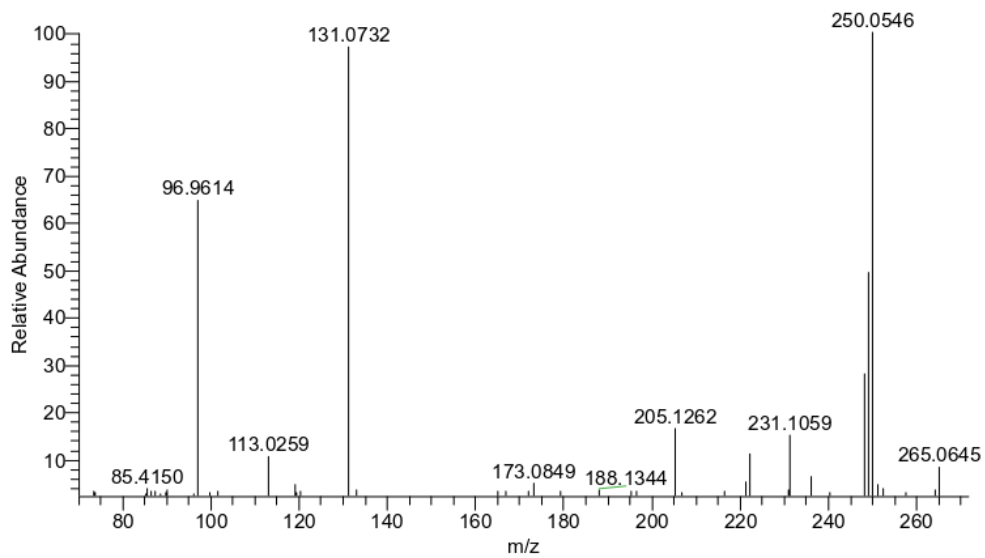
234
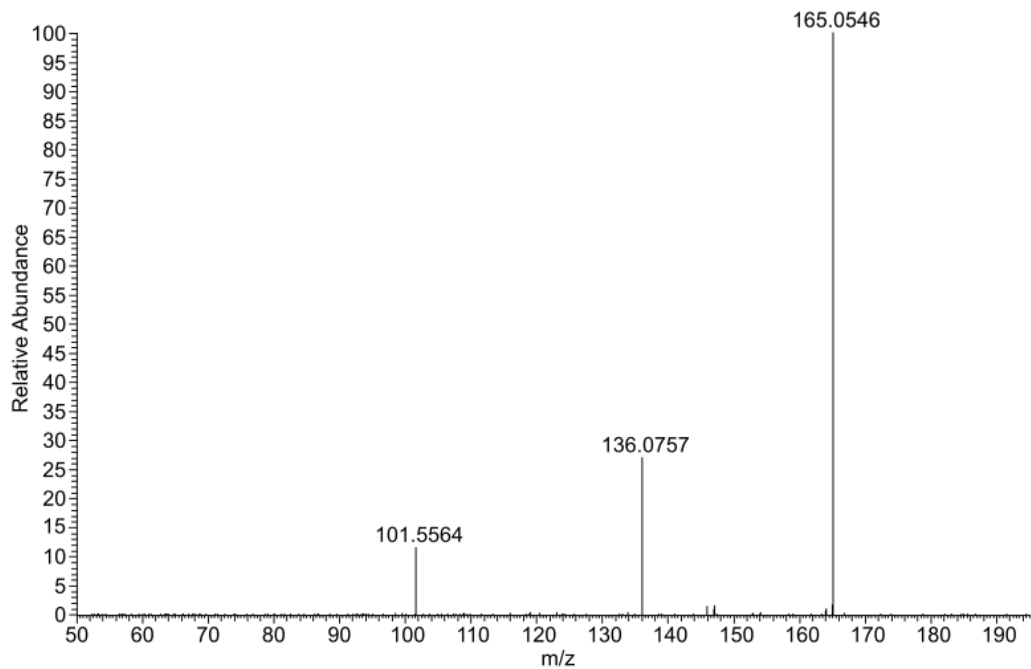
Figure 69: Experimental MS2 spectrum for feature detected by HILIC (+) at 182.0811 *m/z*
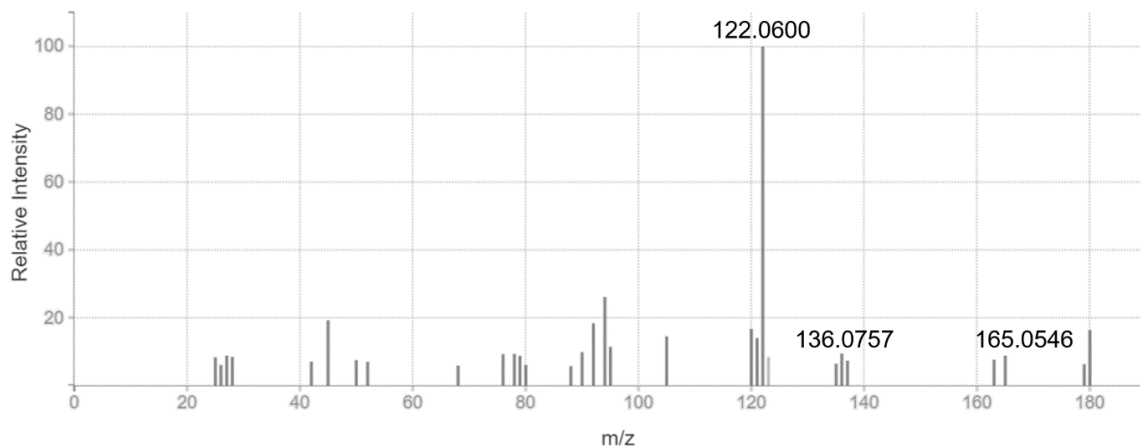
Figure 70: Corresponding database predicted MS^2 spectrum for matched compound at higher CID energy (40 CID), beta-tyrosine

Figure 71: Experimental MS^2 spectrum for feature detected by HILIC (+) at 188.0705 *m/z*
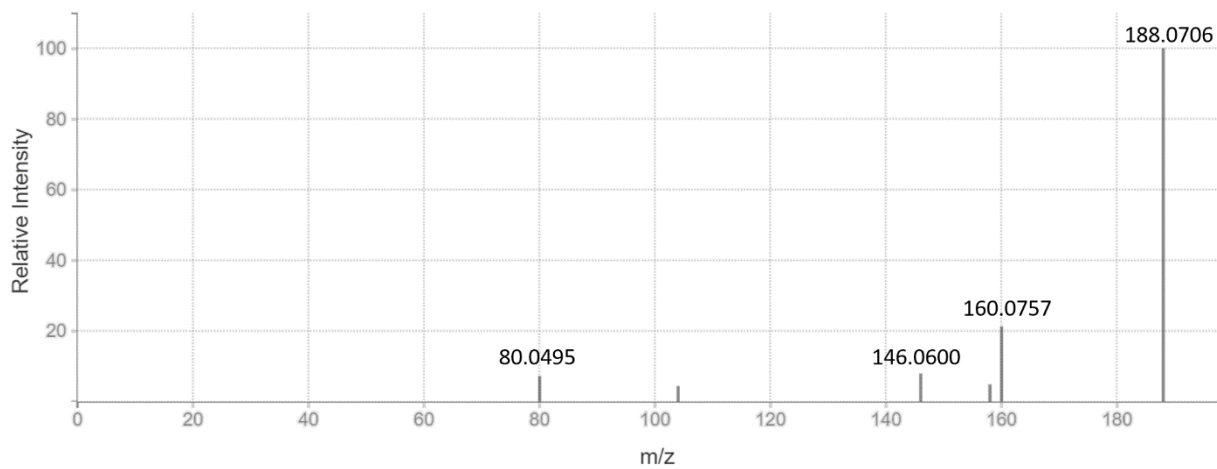
Figure 72: Corresponding database MS^2 spectrum for matched compound, N-(2,5-Dihydroxyphenyl) pyridinium standard (predicted MS^2 at 20 CID energy)
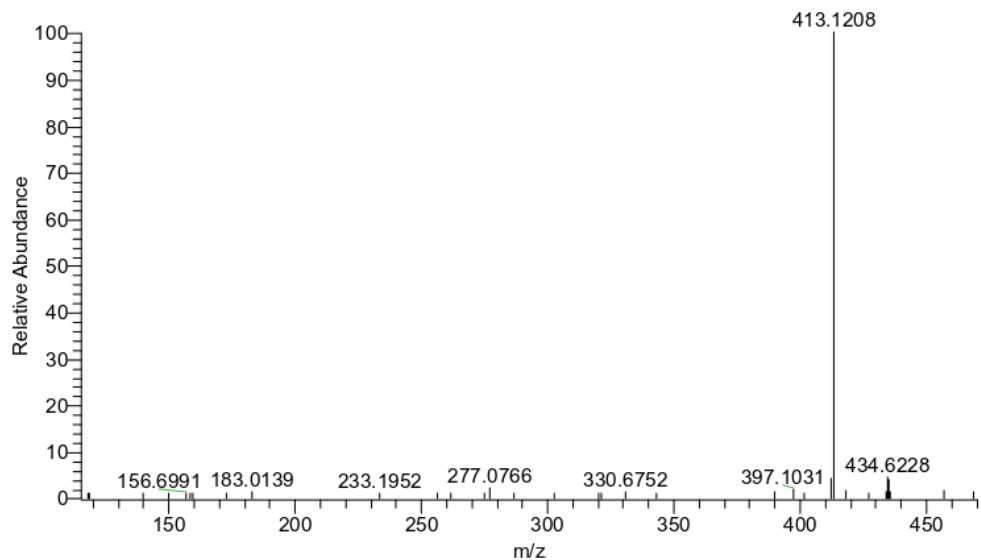
Figure 73: Experimental MS2 spectrum for feature detected by HILIC (+) at 220.1178 *m/z*
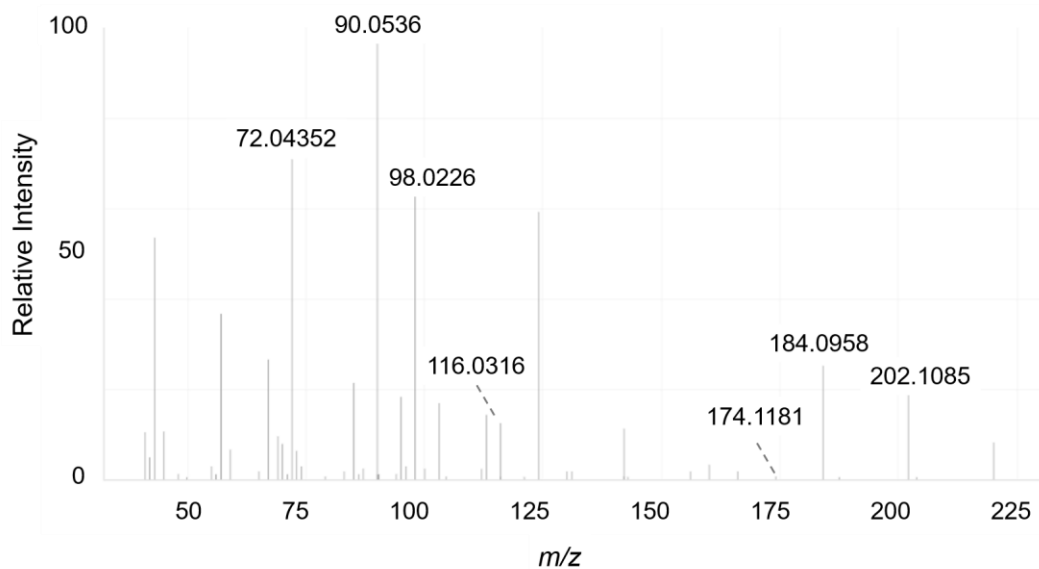
Figure 74: Corresponding database MS^2 spectrum for matched standard, pantothenic acid

Figure 75: Experimental MS^2 spectrum for feature detected by HILIC (-) at 219.1021 *m/z*
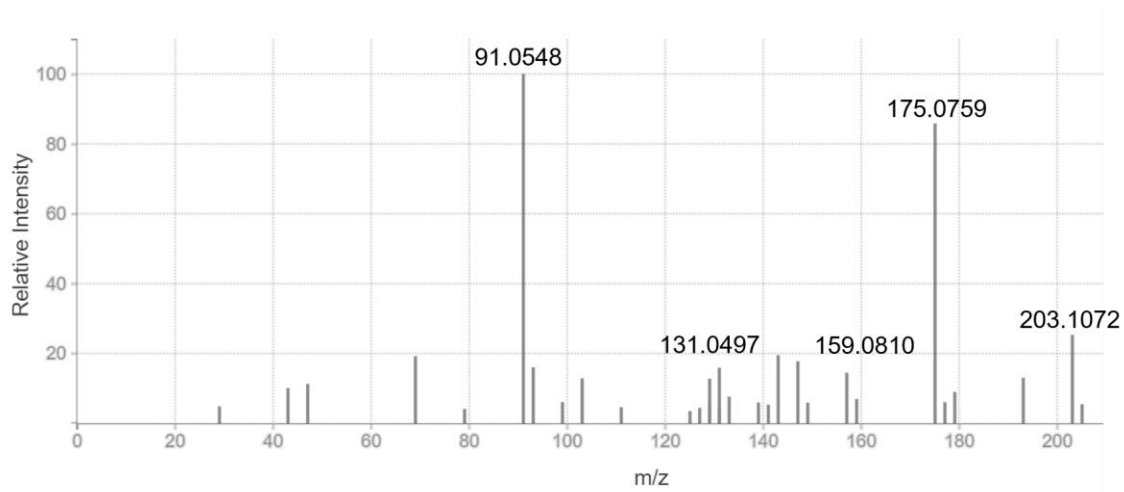
Figure 76: Corresponding database predicted MS^2 spectrum for matched compound, ethyl 2-benzylacetoacetate
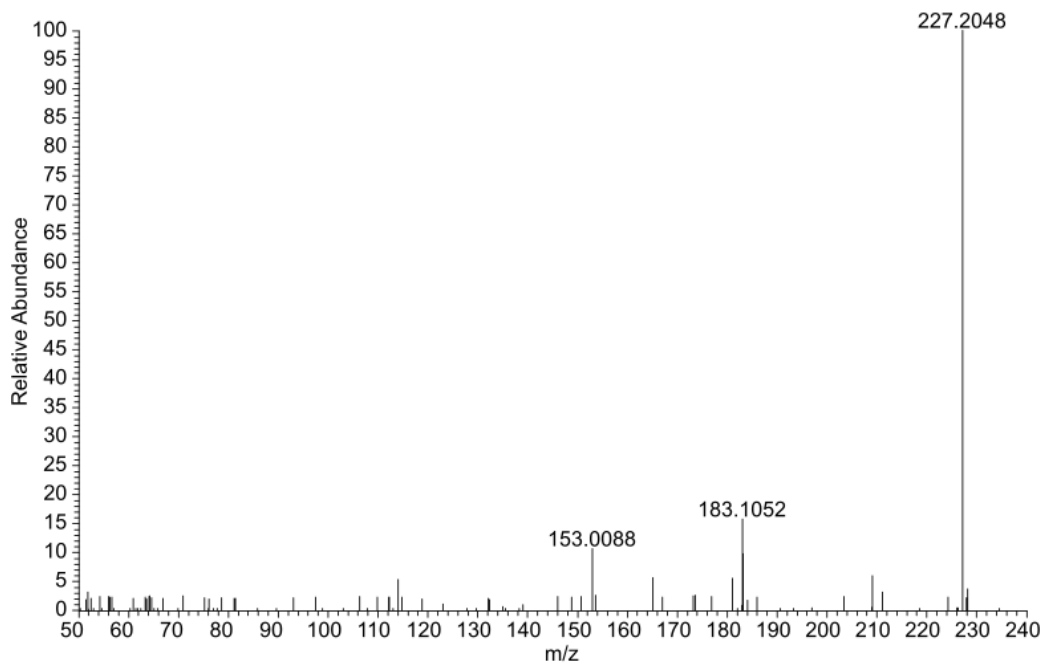
Figure 77: Experimental MS^2 spectrum for feature detected by HILIC (-) at 227.1074 *m/z*
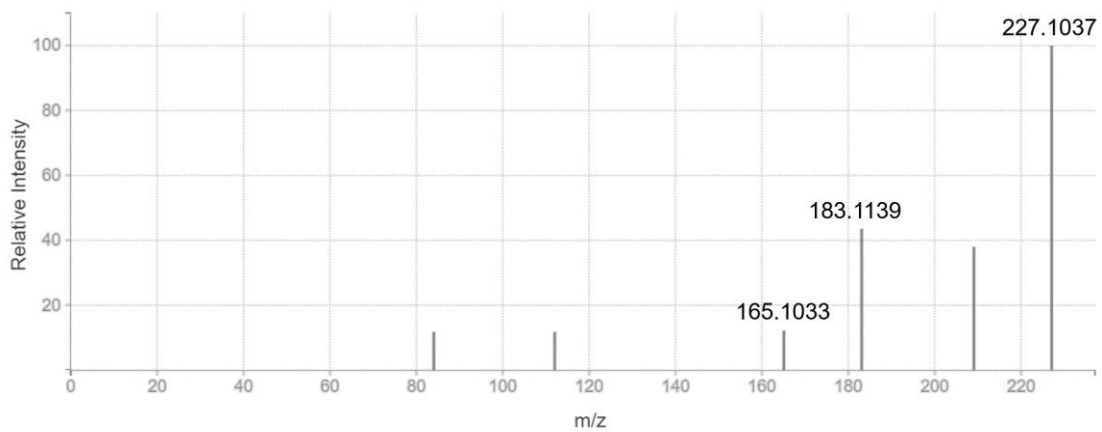
Figure 78: Corresponding database predicted MS^2 spectrum for matched compound, pyroglutamylvaline
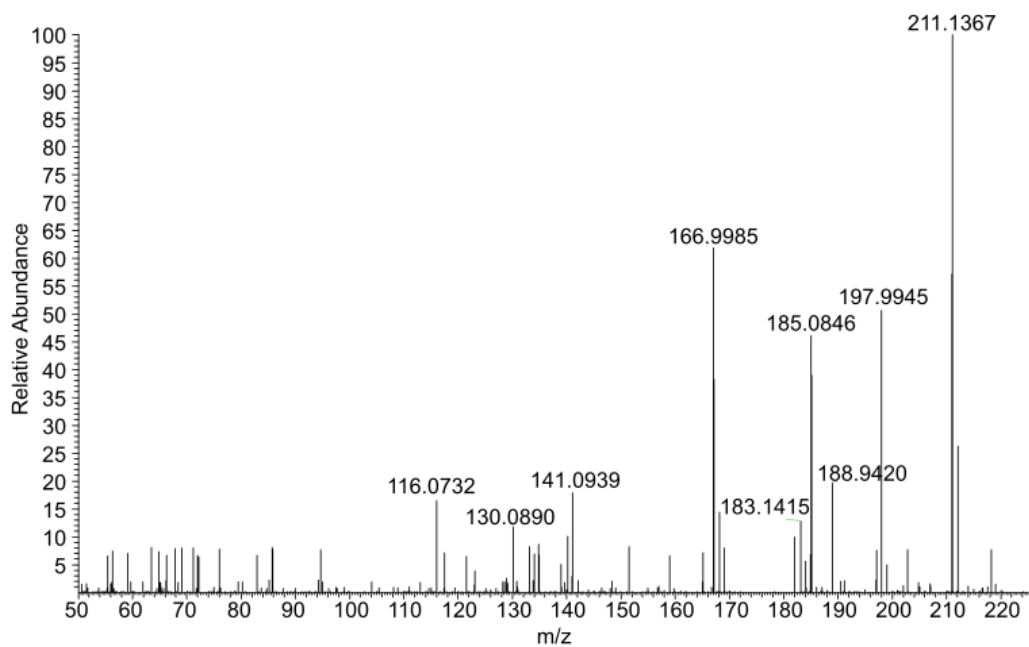
Figure 79: Experimental MS^2 spectrum for feature detected by HILIC (-) at 229.1239 *m/z*
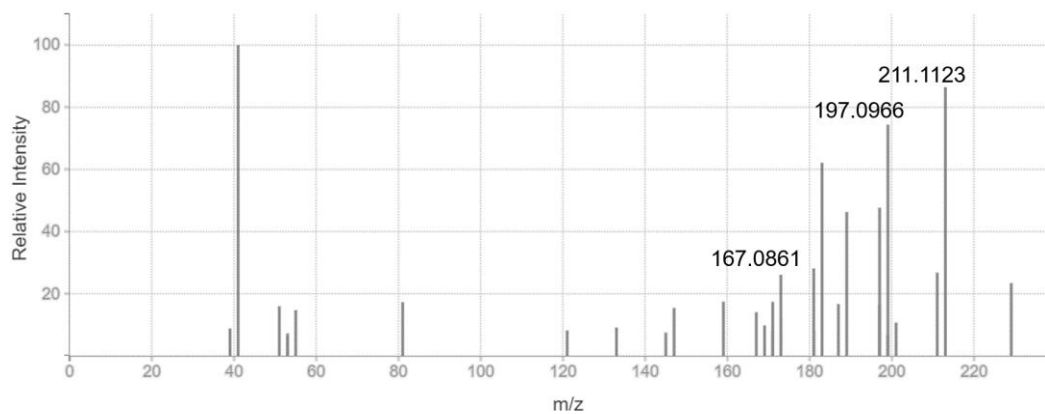
Figure 80: Corresponding database predicted MS^2 spectrum for matched compound, 8,12-Epoxy-4(15),7,11-eudesmatrien-1-one

Figure 81: Experimental MS^2 spectrum for feature detected by HILIC (-) at 293.1442 *m/z*
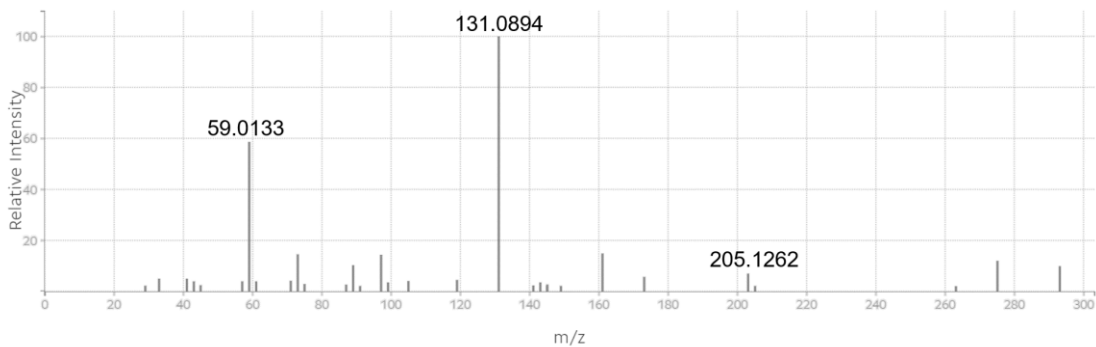
Figure 82: Corresponding database predicted MS^2 spectrum for matched compound, heptyl 1-thiohexopyranoside

Figure 83: Experimental MS^2 spectrum for feature detected by HILIC (-) at 457.1309 *m/z*
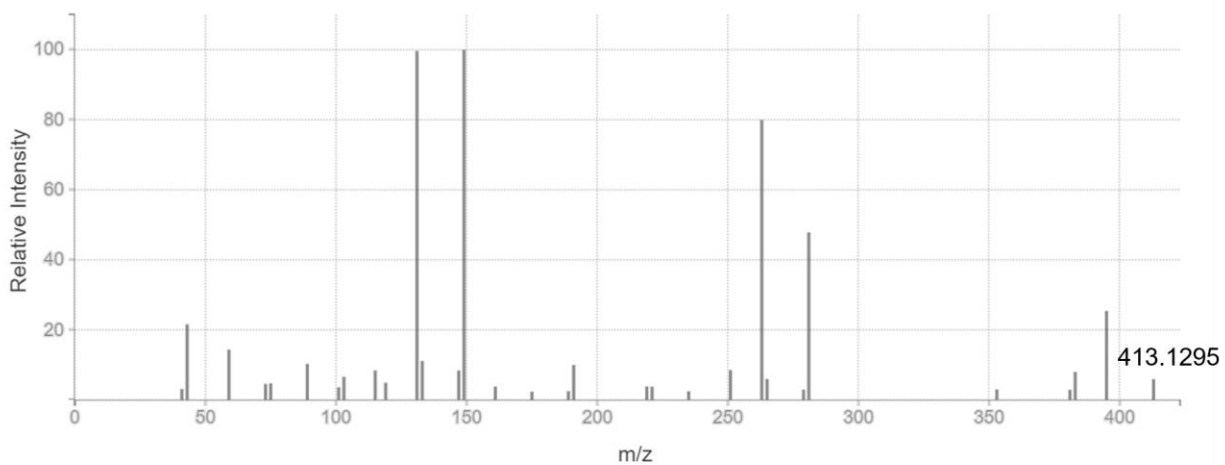
Figure 84: Corresponding database predicted MS^2 spectrum for matched compound, a-L-Arabinofuranosyl-(1->3)-b-D-xylopyranosyl-(1->4)-D-xylose at higher CID energy (40 CID)
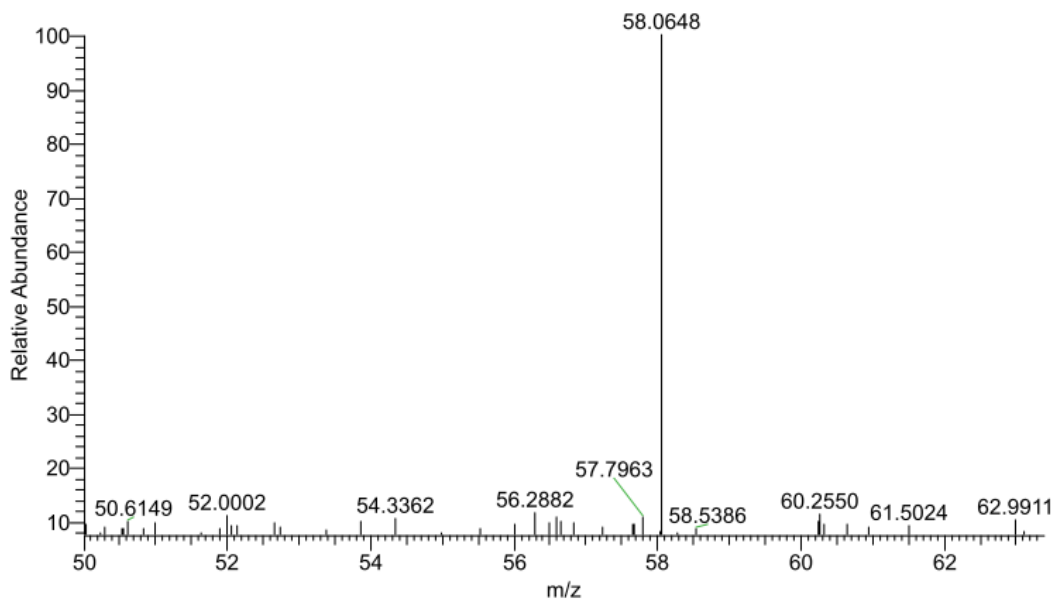
Figure 85: Experimental MS^2 spectrum for feature detected by RP (+) at 60.0444 *m/z*
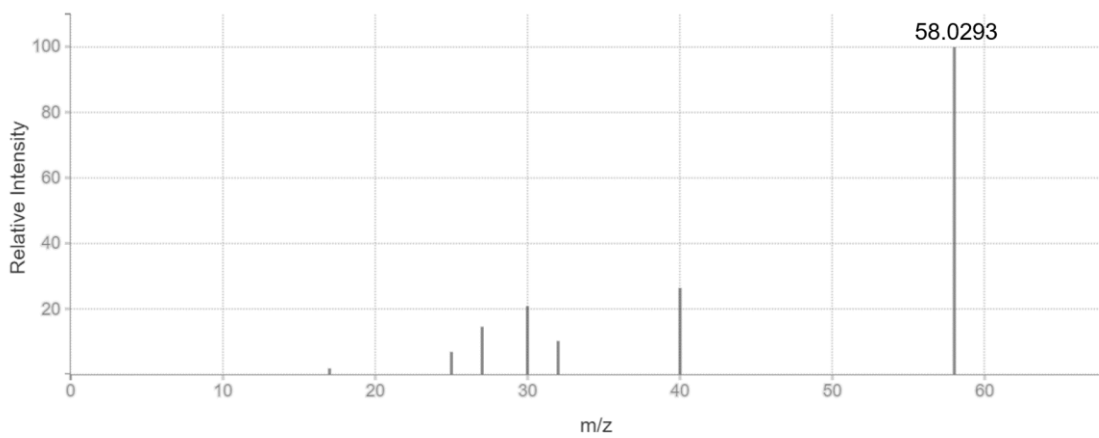
Figure 86: Corresponding database predicted MS^2 spectrum for matched compound, aminoacetaldehyde
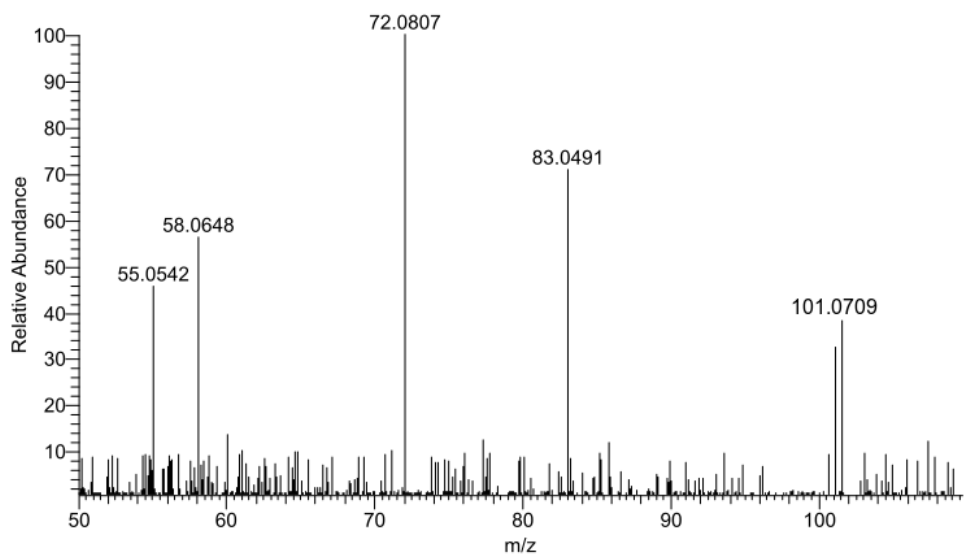
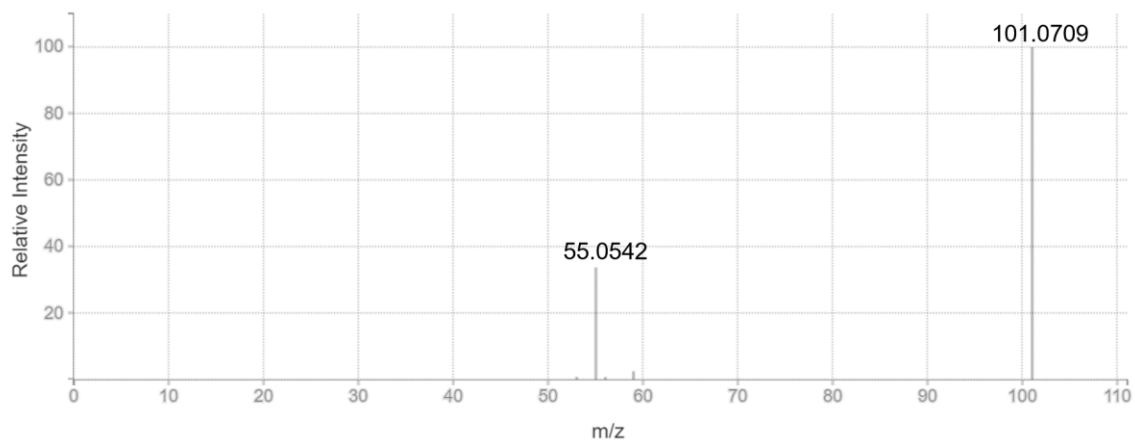**Figure 87:** Experimental MS^2 spectrum for feature detected by RP (+) at 101.0709 *m/z*

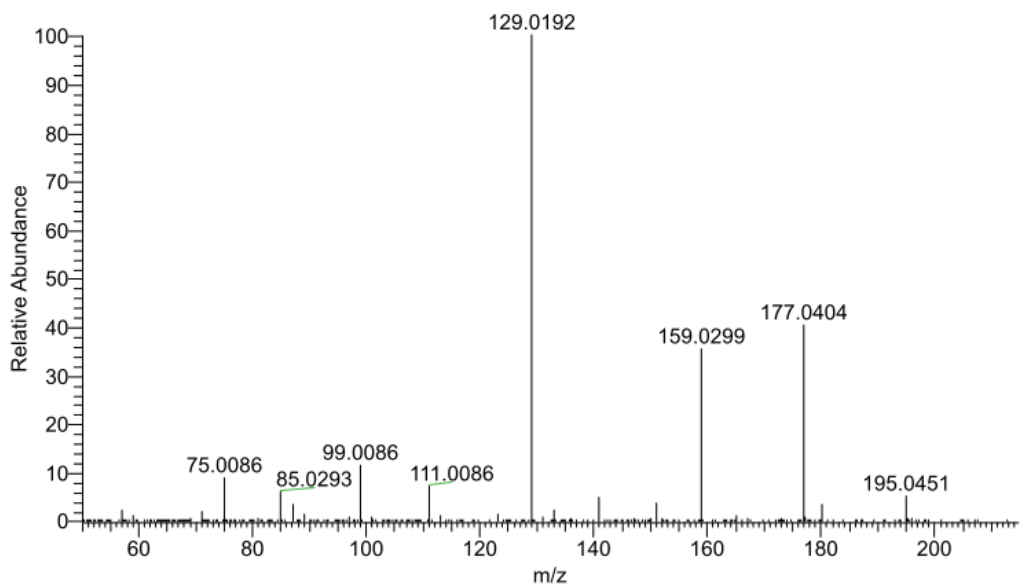**Figure 88:** Corresponding database predicted MS^2 spectrum for matched compound, N-nitroso-pyrrolidine

Figure 89: Experimental MS^2 spectrum for feature detected by RP (+) at 195.0512 *m/z*
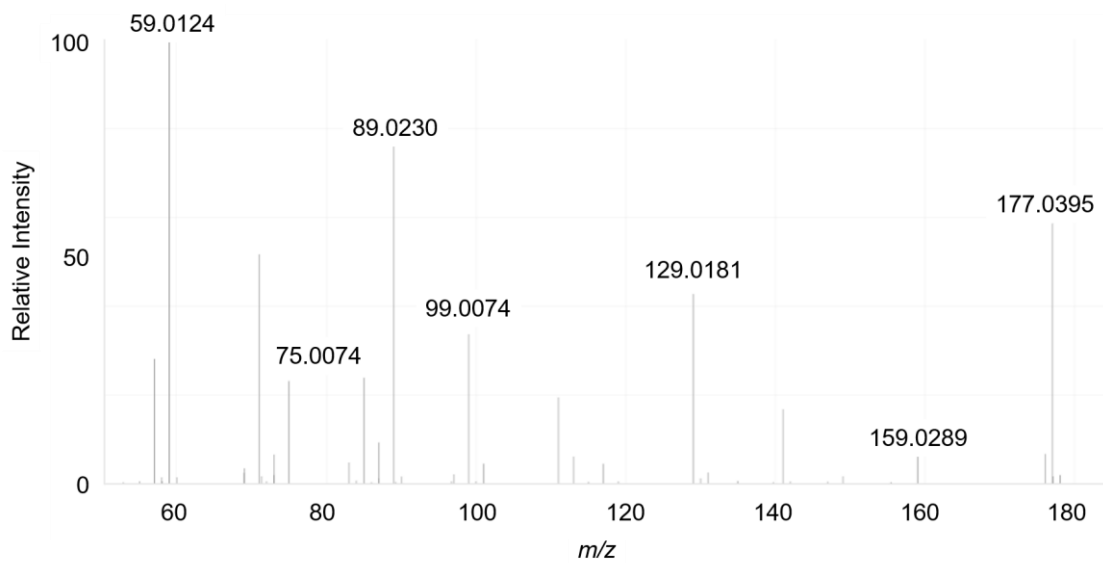

Figure 90: Corresponding database MS^2 spectrum for matched standard, D-gulonic acid

Appendix C: Annotated LMW DOM tables

The attached appendix includes two tables of annotations based on database searching for differentially-abundant features due to polygon type or aboveground vegetation, identified in the study described in Chapter 4 of this dissertation. Databases used are indicated in Chapter 2 and additional information listed includes which tool was used (MZmine or Metabosearch), the unique ID for each feature, the LC/MS condition it was detected by, and the high mass accuracy molecular ion (*m/z*).

VITA

Mallory Paige Ladd was born in Crystal Lake, Illinois where she graduated from Prairie Ridge High School in 2006. She obtained a Bachelor of Science degree in Chemistry from the University of Toledo in Toledo, Ohio with *magna cum laude* honors in 2011, and then worked for two years as a lab manager and field technician in the Alaskan Arctic. In 2013, she was awarded a competitive fellowship to join the Bredesen Center for Interdisciplinary Research and Graduate Education as a graduate student, where she discovered Dr. Robert Hettich's mass spectrometry group and the Next-Generation Ecosystem Experiments project at Oak Ridge National Laboratory. She earned a National Science Foundation Graduate Research Fellowship in 2014 to fund this research and is graduating with a Doctor of Philosophy in Energy Science and Engineering in December 2018. She recently accepted a position as a Research Analyst with the Center for Naval Analyses in Washington, D.C.